PCT/GB 99/02044

The Patent Office

09/720034

INVESTOR IN PEOPLE

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

-registration under the Companies Act does not constitute a new legal entity but merely bjects the company to certain additional company law rules.

Signed

Dated 3 August 1999

Patents Form 1/77

...nts Act 1977
...le 16)

The Patent Office

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

29JUN98 E371361-1 002E23
29 JUN 1998

1. Your reference	IS/CP5710660
2. Patent application number *(The Patent Office will fill in this part)*	29 JUN 1998 **9814006.4**
3. Full name, address and postcode of the or of each applicant *(underline all surnames)*	BIOTICA TECHNOLOGY LIMITED 112 HILLS ROAD CAMBRIDGE CB2 1PH
Patents ADP number *(if you know it)*	
If the applicant is a corporate body, give the country/state of its incorporation	GB 7466824001
4. Title of the invention	POLYKETIDES AND THEIR SYNTHESIS
5. Name of your agent *(if you have one)*	MEWBURN ELLIS
"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*	YORK HOUSE 23 KINGSWAY LONDON WC2B 6HP ✓
Patents ADP number *(if you know it)*	109006

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (if you know it) the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month /year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*
   a) any applicant named in part 3 is not an inventor, or
   b) there is an inventor who is not named as an applicant, or
   c) any named applicant is a corporate body.
   See note (d))

YES

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	0
Description	51
Claim(s)	0
Abstract	0
Drawing(s)	~~7~~ 8.

10. If you are also filing any of the following, state

~~Priority documents~~	~~0~~
Translations of priority documents	0
Statement of inventorship and right	0
Request for preliminary examination	0
Request for substantive examination *(Patents Form 10/77)*	0
Any other documents *(Please specify)*	0

11. I/We request the grant of a patent on the basis of this application.

Signature

Date
29 June 1998

12. Name and daytime telephone number of person to contact in the United Kingdom

IAN STUART

0117 9266411

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed it it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

Polyketides and their Synthesis

The present invention relates to processes and materials (including enzyme systems, nucleic acids, vectors and cultures) for preparing novel polyketides, particularly 12-, 14- and 16-membered ring macrolides, by recombinant synthesis and to the novel polyketides so produced. Polyketide biosynthetic genes or portions of them, which may be derived from different polyketide biosynthetic gene clusters are manipulated to allow the production of specific novel polyketides, such as 12-, 14- and 16-membered macrolides, of predicted structure. The invention is particularly concerned with the replacement of genetic material encoding the natural starter unit with other genes in order to prepare macrolides with preferentially an acetate starter unit; or preferentially a propionate unit; or preferentially with an unusual starter unit, in each case minimising the formation of by-products containing a different starter unit.

Polyketides are a large and structurally diverse class of natural products that includes many compounds possessing antibiotic or other pharmacological properties, such as erythromycin, tetracyclines, rapamycin, avermectin, monensin, epothilones and FK506.

In particular, polyketides are abundantly produced by *Streptomyces* and related actinomycete bacteria. They are synthesised by the repeated stepwise condensation of acylthioesters in a manner analogous to that of fatty acid biosynthesis. The greater structural diversity found among natural polyketides arises from the selection of (usually) acetate or propionate as A starter or "extender" units; and from the differing degree of processing of the -keto group observed after each condensation. Examples of processing steps include reduction to β -hydroxyacyl-, reduction followed by dehydration to 2-enoyl-, and complete reduction to the saturated acylthioester. The stereochemical outcome of these processing steps is also specified for each cycle of chain extension.

The biosynthesis of polyketides is initiated by a group of chain-forming enzymes known as polyketide synthases. Two classes of polyketide synthase (PKS) have been described in actinomycetes. One class, named Type I PKSs, represented by the PKSs for the macrolides erythromycin, oleandomycin, avermectin and rapamycin, consists of a different set or "module" of enzymes for each cycle of polyketide chain extension. For an example see Figure 1 (Cortés, J. *et al.* Nature (1990) 348:176-178; Donadio, S. *et al.* Science (1991) 2523:675-679;

Swan, D.G. *et al.* Mol. Gen. Genet. (1994) 242:358-362; MacNeil, D. J. *et al.* Gene (1992) 115:119-125; Schwecke, T. *et al.* Proc. Natl. Acad. Sci. USA (1995) 92:7839-7843).

The term "extension module" as used herein refers to the set of contiguous domains, from a β-ketoacyl-ACP synthase ("AKS") domain to the next acyl carrier protein ("ACP") domain, which accomplishes one cycle of polyketide chain extension. The term "loading module" is used to refer to any group of contiguous domains which accomplishes the loading of the starter unit onto the PKS and thus renders it available to the KS domain of the first extension module. The length of polyketide formed has been altered, in the case of erythromycin biosynthesis, by specific relocation using genetic engineering of the enzymatic domain of the erythromycin-producing PKS that contains the chain releasing thioesterase/cyclase activity (Cortés *et al.* Science (1995) 268:1487-1489; Kao, C.M. *et al.* J. Am. Chem. Soc. (1995) 117:9105-9106).

In-frame deletion of the DNA encoding part of the ketoreductase domain in module 5 of the erythromycin-producing PKS (also known as 6-deoxyerythronolide B synthase, DEBS) has been shown to lead to the formation of erythromycin analogues 5,6-dideoxy-3-α-mycarosyl-5-

oxoerythronolide B, 5,6-dideoxy-5-oxoerythronolide B and 5,6-dideoxy, 6 β-epoxy-5-oxoerythronolide B (Donadio, S. *et al.* Science (1991) 252:675-679). Likewise, alteration of active site residues in the enoylreductase domain of module 4 in DEBS, by genetic engineering of the corresponding PKS-encoding DNA and its introduction into *Saccharopolyspora erythraea*, led to the production of 6,7-anhydroerythromycin C (Donadio, S. *et al.* Proc Natl. Acad. Sci. USA (1993) 90:7119-7123).

International Patent Application number WO 93/13663 describes additional types of genetic manipulation of the DEBS genes that are capable of producing altered polyketides. However many such attempts are reported to have been unproductive (Hutchinson, C. R. and Fujii, I. Annu. Rev. Microbiol. (1995) 49:201-238, at p. 231). The complete DNA sequence of the genes from *Streptomyces hygroscopicus* that encode the modular Type I PKS governing the biosynthesis of the macrocyclic immunosuppressant polyketide rapamycin has been disclosed (Schwecke, T. *et al.* (1995) Proc. Natl. Acad. Sci. USA 92:7839-7843). The DNA sequence is deposited in the EMBL/Genbank Database under the accession number X86780.

The second class of PKS, named Type II PKSs, is represented by the synthases for aromatic compounds. Type II PKSs contain only a single set of enzymatic

activities for chain extension and these are re-used as appropriate in successive cycles (Bibb, M. J. *et al.* EMBO J. (1989) 8:2727-2736; Sherman, D. H. *et al.* EMBO J. (1989) 8:2717-2725; Fernandez-Moreno, M.A. *et al.* J. Biol. Chem. (1992) 267:19278-19290). The "extender" units for the Type II pKSs are usually acetate units, and the presence of specific cyclases dictates the preferred pathway for cyclisation of the completed chain into an aromatic product (Hutchinson, C. R. and Fujii, I. Annu. Rev. Microbiol. (1995) 49:201-238). Hybrid polyketides have been obtained by the introduction of clones Type II PKS gene-containing DNA into another strain containing a different Type II PKS gene cluster, for example by introduction of DNA derived from the gene cluster for actinorhodin, a blue-pigmented polyketide from *Streptomyces coelicolor*, into an anthraquinone polyketide-producing strain of *Streptomyces galileus* (Bartel, P. L. *et al.* J. Bacteriol. (1990) 172:4816-4826).

The minimal number of domains required for polyketide chain extension on a Type II PKS when expressed in a *Streptomyces coelicolor* host cell (the "minimal PKS") has been defined for example in International Patent Application Number WO 95/08548 as containing the following three polypeptides which are

products of the *act I* genes: first KS; secondly a polypeptide termed the CLF with end-to-end amino acid sequence similarity to the KS but in which the essential active site residue of the KS, namely a cysteine residue, is substituted either by a glutamine residue, or in the case of the PKS for a spore pigment such as the *whiE* gene product (Chater, K. F. and Davis, N. K. Mol. Microbiol. (1990) 4:1679-1691) by a glutamic acid residue (Figure 2); and finally an ACP. The CLF has been stated for example in International Patent Application Number WO 95/08548 to be a factor that determines the chain length of the polyketide chain that is produced by the minimal PKS. However it has been found (Shen, B. *et al.* J. Am. Chem. Soc. (1995) 117:6811-6821) that when the CLF for the octaketide actinorhodin is used to replace the CLF for the decaketide tetracenomycin in host cells of *Streptomyces glaucescens*, the polyketide product is not found to be altered from a decaketide to an octaketide, so the exact role of the CLF remains unclear. An alternative nomenclature has been proposed in which KS is designated KSα and CLF is designated KSβ, to reflect this lack of knowledge (Meurer, G. *et al.* Chemistry and Biology (1997) 4:433-443). The mechanism by which acetate starter units and acetate extender units are loaded onto the Type II PKS is not known, but it is

speculated that the malonyl-CoA: ACP acyltransferase of the fatty acid synthase of the host cell can fulfil the same function for the Type II PKS (Revill, W. P. *et al.* J. Bacteriol. (1995) 177:3946-3952).

International Patent Application Number WO 95/08548 describes the replacement of actinorhodin PKS genes by heterologous DNA from other Type II PKS gene clusters, to obtain hybrid polyketides. The same International Patent Application WO 95/08548 describes the construction of a strain of *Streptomyces coelicolor* which substantially lacks the native gene cluster for actinorhodin, and the use in that strain of a plasmid vector pRM5 derived from the low-copy number vector SCP2* isolated from *Streptomyces coelicolor* (Bibb, M. J. and Hopwood, D. A. J. Gen. Microbiol. (1981) 126:427) and in which heterologous PKS-encoding DNA may be expressed under the control of the divergent *act I/ act* III promoter region of the actinorhodin gene cluster (Fernandez-Moreno, M.A. *et al.* J. Biol. Chem. (1992) 267:19278-19290). The plasmid pRM5 also contains DNA from the actinorhodin biosynthetic gene cluster encoding the gene for a specific activator protein, ActII-orf4. The Act II-orf4 protein is required for transcription of the genes placed under the control of the actI/ act II bidirectional promoter and activates gene expression during the

transition from growth to stationary phase in the vegetative mycelium (Hallam, S. E. *et al.* Gene (1988) 74:305-320).

Type II clusters in *Streptomyces* are known to be activated by pathway-specific activator genes (Narva, K. E. and Feitelson, J. S. J. Bacteriol. (1990) 172:326-333; Stutzman-Engwall, K. J. *et al.* J. Bacteriol. (1992) 174:144-154; Fernandez-Moreno, M.A. *et al.* Cell (1991) 66:769-780; Takano, E. *et al.* Mol. Microbiol. (1992) 6:2797-2804; Takano, E. *et al.* Mol. Microbiol. (1992) 7:837-845), The DnrI gene product complements a mutation in the *actII*-orf4 gene of *S. coelicolor*, implying that DnrI and ActII-orf4 proteins act on similar targets. A gene (*srmR*) has been described (EP 0 524 832 A2) that is located near the Type I PKS gene cluster for the macrolide polyketide spiramycin. This gene specifically activates the production of the macrolide antibiotic spiramycin, but no other exampples have been found of such a gene. Also, no homologues of the ActII-orf4/DnrI/RedD family of activators have been described that act on Type I PKS genes.

Although large numbers of therapeutically important polyketides have been identified, there remains a need to obtain novel polyketides that have enhanced properties or possess completely novel bioactivity. The complex

polyketides produced by Type I PKSs are particularly valuable, in that they include compounds with known utility as anthelminthics, insecticides, immunosuppressants, antifungal or antibacterial agents. Because of their structural complexity, such novel polyketides are not readily obtainable by total chemical synthesis, or by chemical modifications of known polyketides.

There is also a need to develop reliable and specific ways of deploying individual modules in practice so that all, or a large fraction, of hybrid PKS genes that are constructed, are viable and produce the desired polyketide product.

Pending International Patent Application number PCT/GB97/01818 discloses that a PKS gene assembly (particularly of Type I) encodes a loading module which is followed by at least one extension module. Thus Figure 1 shows the organisation of the DEBS genes. The first open reading frame encodes the first multi-enzyme or cassette (DEBS 1) which consists of three modules: the loading module (ery-load) and two extension modules (modules 1 and 2). The loading module comprises an acyltransferase and an acyl carrier protein. This may be contrasted with Fig. 1 of WO 93/13663 (referred to above). This shows ORF1 to consist of only two modules,

the first of which is in fact both the loading module and the first extension module.

PCT/GB97/01818 describes in general terms the production of a hybrid PKS gene assembly comprising a loading module and at least one extension module.

PCT/GB97/01818 also describes (see also Marsden, A. F. A. *et al*. Science (1998) 279:199-202) construction of a hybrid PKS gene assembly by grafting the wide-specificity loading module for the avermectin-producing polyketide synthase onto the first multienzyme component (DEBS 1) for the erythromycin PKS in place of the normal loading module. Certain novel polyketides can be prepared using the hybrid PKS gene assembly, as described for example in pending International Patent Application number (PCT/GB97/01810). Patent Application PCT/GB97/01818 further describes the construction of a hybrid PKS gene assembly by grafting the loading module for the rapamycin-producing polyketide synthase onto the first multienzyme component (DEBS 1) for the erythromycin PKS in place of the normal loading module. The loading module of the rapamycin PKS differs from the loading modules of DEBS and the avermectin PKS in that it comprises a CoA ligase domain, an enoylreductase ("ER") domain and an ACP, so that suitable organic acids including the natural starter unit 3,4-

dihydroxycyclohexane carboxylic acid may be activated in situ on the PKS loading domain, and with or without reduction by the ER domain transferred to the ACP for intramolecular loading of the KS of extension module 1 (Schwecke, T. *et al.* Proc. Natl. Acad. Sci. USA (1995) 92:7839-7843).

The DNA sequences have been disclosed for several Type I PKS gene clusters that govern the production of 16-membered macrolide polyketides, including the tylosin PKS from *Streptomyces fradiae* (EP 0 791 655 A2), the niddamycin PKS from *Streptomyces caelestis* (Kavakas, S. J. *et al.* J. Bacteriol. (1998) 179:7515-7522) and the spiramycin PKS from *Streptomyces ambofaciens* (EP 0791 655 A2). All of these gene sequences have in common that they show the loading module of the PKS to differ from the loading module of DEBS and of the avermectin PKS in that they consist of a domain resembling the KS domains of the extension modules, an AT domain and an ACP (Figure 3). The additional N-terminal KS-like domain has been named KSq because it differs in each case from an extension KS by the specific replacement of the active site cysteine residue essential for β-ketoacyl-ACP synthase activity by a glutamine (Q in single letter notation) residue. The function of the KSq domain is unknown (Kavakas, S. J. *et al.* J. Bacteriol. (1998)

179:7515-7522), but its presence in these PKSs for 16-membered macrolides is surprising because the starter units of tylosin, niddamycin and spiramycin appear to be propionate, acetate and acetate respectively, that is, the same type of starter unit as in DEBS. The AT adjacent to the KSq domain is named here the ATq domain.

When the entire loading module of the tylosin PKS was used to replace the analogous loading module in the spiramycin PKS in *S. ambofaciens* (Kuhstoss *et al.* Gene (1996) 183:231-236), the nature of the starting unit was stated to be altered from acetate to propionate. Since the role of the KSq domain was not understood, no specific disclosure was made that revealed either the importance of the KSq domain, or the possible utility of these KSq-containing loading modules in ensuring the purity of the polyketide product in respect of the starter unit, even at high levels of macrolide production. The interpretation for their results was stated as: "Therefore we believe that the experiments described here provide strong experimental support for the hypothesis that the AT domains in Type I PKS systems select the appropriate substrate at each step in synthesis" (Kuhstoss *et al.* Gene (1996) 183:231-236, at p. 235). These authors noted the analogy with the CLF protein in Type II PKS systems and that the latter

protein is thought to be involved in determining the chain length. They state: "KSq may serve a similar function, although it is unclear why such a function would be necessary in the synthesis of these 16-membered polyketides when it is not needed for the synthesis of other complex polyketides such as 6-DEB or rapamycin. In any case the KSq is unlikely to be involved in substrate choice at each step of synthesis." (Kuhstoss *et al.* Gene (1996) 183:231-236).

It has been shown that when genetic engineering is used to remove the loading module of DEBS, the resulting truncated DEBS in *S. erythraea* continues to produce low levels of erythromycins containing a propionate starter unit (Pereda, A. *et al.* Microbiology (1995) 144:543-553). The same publication shows that when in this truncated DEBS the methylmalonyl-CoA -specific AT of extension module 1 was replaced by a malonyl-CoA-specific AT from an extension module of the rapamycin PKS, the products were also low levels of erythromycins containing a propionate starter unit, demonstrating that the origin of the starter units is not decarboxylation of the (methyl)malonyl groups loaded onto the enzyme by the AT of module 1, but from direct acylation of the KS of extension module 1 by propionyl-CoA. This is in contrast to a previous report, using partially purified DEBS1+TE,

a truncated bimodular PKS derived from DEBS (Kao, C. M. *et al.* J. Am. Chem. Soc. (1995) 117:9105-9106) and functionally equivalent to DEBS1-TE (Brown, M. J. B. *et al.*, J. Chem. Soc. Chem. Commun. (1995) 1517-1518; Cortés, J. *et al.* Science (1991) 2523:675-679), which stated that the origins of the starter units for DEBS can include methylmalonate units which are loaded onto module 1 and are decarboxylated by the KS of module 1 (Pieper, R. *et al.* Biochemistry (1997) 36:1846-1851). It has now been found that when the DEBS1-TE protein is fully purified from extracts of recombinant *S. erythraea* it contains no such specific decarboxylase activity (Weissmann, K. et al. (1998) Biochemistry, in the press), further confirming that starter units do not in fact arise from decarboxylation of extension units mediated by the KS of extension module 1 .

It is known that the DEBS loading module has a slightly broader specificity than propionate only, and in particular acetate starter units are used both in vitro and in vivo, when the PKS containing this loading module is part of a PKS that is expressed either in *S. erythraea* the natural host for erythromycin production (see for example Cortés, J. *et al.* Science (1995) 268:1487-1489), or in a heterologous host such as *S. coelicolor* (Kao, C. M. *et al.* J. Am. Chem. Soc. (1994) 116:11612-11613;

Brown, M. J. B. *et al.* J. Chem. Soc. Chem. Commun. (1995) 1517-1519). In vitro experiments using purified DEBS1-TE have demonstrated that propionyl-CoA and acetyl-CoA are alternative substrates that efficiently supply propionate and acetate units respectively to the loading module (Wiessmann, K. E. H. *et al.* Chemistry and Biology (1995) 2:583-589; Pieper, R. *et al.* J. Am. Chem. Soc. (1995) 117:11373-11374). The outcome of the competition between acetate and propionate starter units is influenced by the respective intracellular concentrations of propionyl-CoA and acetyl-CoA prevailing in the host cell used (see for example Kao, C. M. *et al.* Science (1994) 265:509-512; Pereda, A. *et al.* Microbiology (1995) 144:543-553). It is also determined by the level of expression of the host PKS, so that as disclosed for example in Pending International Patent Application number PCT/GB97/01818, when recombinant DEBS or another hybrid PKS containing the DEBS loading module is over-expressed in *S. erythraea*, the products are generally mixtures whose components differ only in the presence of either an acetate or a propionate starter unit.

There is a need to develop reliable and methods for avoiding the formation of mixtures of polyketides with both acetate and propionate starter units, and to allow the specific incorporation of unusual starter units. It

has now been found, surprisingly, that the role of the loading domains in the PKSs for the 16-membered macrolides tylosin, niddamycin and spiramycin is different from that of the loading domains of the avermectin PKS and of DEBS. It has been realised that the KSq domain of the tylosin PKS and the associated AT domain, which is named here ATq, together are responsible for the highly specific production of propionate starter units because the ATq is specific for the loading of methylmalonyl-CoA and not propionyl-CoA as previously thought; and the KSq is responsible for the highly specific decarboxylation of the enzyme-bound methylmalonate unit to form propionate unit attached to the ACP domain of the loading module and appropriately placed to be transferred to the KS of extension module 1 for the initiation of chain extension. In a like manner the ATq of the spiramycin and niddamycin PKSs, and the adjacent KSq, are responsible for the specific loading of malonate units rather than acetate units as previously believed, and for their subsequent specific decarboxylation to provide acetate starter units for polyketide chain extension.

It has also now been found here that not only the PKSs for the above-mentioned 16-membered macrolides, but also the PKSs for certain 14-membered macrolides

particularly the oleandomycin PKS from *Streptomyces antibioticus* (Figure 4) and also the PKSs for certain polyether ionophore polyketides particularly the putative monensin PKS from *Streptomyces cinnamonensis* (Figure 4), possess a loading domain comprising a KSq domain, an ATq domain, and an ACP. In Figure 4 is shown a sequence alignment of the KSq domains and of the adjacent linked ATq domains that have been identified, showing the conserved active site glutamine (Q) residue in the KSq domains, and an arginine residue which is conserved in all extension AT domains and is also completely conserved in ATq domains. This residue is characteristically not arginine in the AT domains of either DEBS or of the avermectin PKS loading modules, where the substrate for the AT is a non-carboxylated acyl-CoA ester (Haydock, S. F. *et al.* FEBS Letters (1995) 374:246-248) . The abbreviation ATq is used here to simply to distinguish the AT domains found immediately C-terminal of Ksq from extension ATs, and the label has no other significance.

In one aspect the invention provides a PKS multienzyme or part thereof, or nucleic acid (generally DNA) encoding it, said multienzyme or part comprising a loading module and a plurality of extension modules, wherein

(a) the loading module is adapted to load a malonyl

or substituted malonyl residue and then to effect decarboxylation of the loaded residue to provide an acetyl or substituted acetyl (which term encompasses propionyl) residue for transfer to an extension module; and

(b) the extension modules, or at least one thereof (preferably at least the one adjacent the loading module), are not naturally associated with a loading module that effects decarboxylation of an optionally substituted malonyl residue.

Generally the loading module will also include an ACP (acyl carrier protein) domain.

Preferably the decarboxylating functionality of the loading module is provided by a KS (ketosynthase)-type domain. Suitably this differs from a KS of a conventional extension module by possessing a glutamine residue in place of the essential cysteine residue in the active site. It is termed Ksq. It may be "natural" or genetically engineered, e.g. resulting from site-directed mutagenesis of nucleic acid encoding a different KS such as a KS of an extension module.

Alternatively the decarboxylating functionality can be provided by a CLF-type domain of the general type occuring in Type II PKS systems.

Preferably the loading functionality is provided by

an AT (acyltransferase)-type domain which resembles an AT domain of a conventional extension module in having an arginine residue in the active site, which is not the case with the AT domains of loader modules which load acetate or propionate, e.g. in DEBS or avermectin PKS systems. It may be termed Atq. Once again, it may be "natural" or genetically engineered, e.g. by mutagenesis of an AT of an extension module.

Usually the loading module will be of the form:

Ksq-ATq-ACP

where ACP is acyl carrier protein.

In another aspect the invention provides a method of synthesising a polyketide having substantially exclusively a desired starter unit by providing a PKS multienzyme incorporating a loading module as defined above which specifically provides the desired starter unit. This may comprise providing nucleic acid encoding the multienzyme and introducing it into an organism where it can be expressed.

In further aspects the invention provides vectors and transformant organisms and cultures containing nucleic acid encoding the multienzyme. A preferred embodiment is a culture which produces a polyketide having a desired starter unit characterised by the substantial absence of polyketides with different starter

units. Thus, for example, erythromycin can be produced substantially free from analogues resulting from the incorporation of acetate starter units in place of propionate.

Preferably the hybrid PKS encodes a loading module and from 2 to 7 extension modules and a chain terminating enzyme (generally a thioesterase).

It is particularly useful to provide a loading module of the type KSq - ATq-ACP for a PKS gene assembly which produces a 12-, 14- or 16-membered macrolide in order to prepare a 12-, 14- or 16-membered macrolide which contains exclusively or almost exclusively an acetate starter unit, even when such PKS gene assembly is expressed at high levels in an actinomycete host cell. Particularly suitable PKSs for this purpose are the components of PKSs for the biosynthesis of erythromycin, methymycin, oleandomycin, tylosin, spiramycin, midecamycin, and niddamycin for all of which the gene and modular organisation is known at least in part. Particularly suitable sources of the genes encoding a loading module of the type KSq - ATq-ACP are the loading modules of oleandomycin, spiramycin, niddamycin, methymycin and monensin which are specific for the loading of malonate units which are then decarboxylated to acetate starter units.

Similarly it is particularly useful to provide a loading module of the type KSq - ATq-ACP for a PKS gene assembly which produces the macrolides rifamycin, avermectin, rapamycin, immunomycin and FK506 whose loading modules possess an unusual specificity (and for all of which the gene and modular organisation is known at least in part) in order to prepare such macrolides containing exclusively or almost exclusively an acetate starter unit, even when such PKS gene assembly is expressed at high levels in an actinomycete host cell. Particularly suitable sources of the genes encoding a loading module of the type KSq - ATq-ACP are the loading modules of oleandomycin, spiramycin, niddamycin, methymycin and monensin which are specific for the loading of malonate units which are decarboxylated to acetate starter units.

It is similarly useful to provide a loading module of the type KSq - ATq-ACP for a PKS gene assembly which produces a 12-, 14- or 16-membered macrolide in order to prepare a 12-, 14- or 16-membered macrolide which contains exclusively or almost exclusively a propionate starter unit, even when such PKS gene assembly is expressed at high levels in an actinomycete host cell. Particularly suitable PKSs for this purpose are the components of PKSs for the biosynthesis of erythromycin,

methymycin, oleandomycin, tylosin, spiramycin, midecamycin, and niddamycin for all of which the gene and modular organisation is known at least in part. A particularly suitable source of the genes encoding a loading module of the type KSq-ATq-ACP is the loading module of tylosin which is specific for the loading of methylmalonate units which are decarboxylated to propionate starter units.

Similarly it is particularly useful to provide a loading module of the type KSq - ATq-ACP for a PKS gene assembly which produces the macrolides rifamycin, avermectin, rapamycin, immunomycin and FK506 whose loading modules possess an unusual specificity (and for all of which the gene and modular organisation is known at least in part) in order to prepare such macrolides containing exclusively or almost exclusively a propionate starter unit, even when such PKS gene assembly is expressed at high levels in an actinomycete host cell. A particularly suitable source of the genes encoding a loading module of the type KSq - ATq-ACP is the loading module of tylosin which is specific for the loading of methylmalonate units which are decarboxylated to propionate starter units.

In the loading module of the type KSq - ATq-ACP the domains or portions of them may be derived from the same

or from different sources, and comprise either natural or engineered domains. For example the ATq domain can be replaced by an AT domain derived from any extension module of a Type I PKS, having specificity either for loading of malonate units or for loading of methylmalonate units respectively, so long as the KSq domain is chosen to have a matching specificity towards either methylmalonate or malonate units respectively.

Alternatively, the KSq domain in the loading module provided of the type KSq - ATq-ACP may be substituted by the CLF polypeptide of a Type II PKS. It is now apparent that in contrast to its previous identification as a factor uniquely determining chain length, the CLF, in addition to any other activities that it may possess, is the analogue of the KSq domain and can act as a decarboxylase towards bound malonate units.

The loading module of the type KSq - ATq-ACP may be linked to a hybrid PKS produced for example as in PCT/GB97/01818 and PCT/GB97/01810. It is particularly useful to link such a loading module to gene assemblies that encode hybrid PKSs that produce novel derivatives of 14-membered macrolides as described for example in PCT/GB97/01818 and PCT/GB97/01810.

The invention further provides such PKS assemblies furnished with a loading module of the type KSq - ATq-

ACP, vectors containing such assemblies, and transformant organisms that can express them. Transformant organisms may harbour recombinant plasmids, or the plasmids may integrate. A plasmid with an *int* sequence will integrate into a specific attachment site (*att*) of the host's chromosome. Transformant organisms may be capable of modifying the initial products, eg by carrying out all or some of the biosynthetic modifications normal in the production of erythromycins (as shown in Figure 5) and for other polyketides. Use may be made of mutant organisms such that some of the normal pathways are blocked, e.g. to produce products withut one or more "natural" hydroxy-groups or sugar groups. The invention further provides novel polyketides as producible, directly or indirectly, by transformant organisms. This includes polyketides which have undergone enzymatic modification.

In a further aspect the invention provides both previously-obtained polyketides and novel polyketides in a purer form with respect to the nature of the starter unit, than was hitherto possible. These include 12- , 14- and 16-membered ring macrolides which are either "natural" or may differ from the corresponding "natural" compound:

a) in the oxidation state of one or more of the ketide units (ie selection of alternatives from the group: -CO-, -CH(OH)-, alkene -CH-, and $-CH_2-$) where the stereochemistry of any -CH(OH)- is also independently selectable;

b) in the absence of a "natural" methyl side-chain; or

c) in the stereochemistry of "natural" methyl; and/or ring substituents other than methyl.

It is also possible to prepare derivatives of 12-, 14- and 16-membered ring macrolides having the differences from the natural product identified in two or more of items a) to c) above.

Derivatives of any of the afore-mentioned polyketides which have undergone further processing by non-PKS enzymes, eg one or more of hydroxylation, epoxidation, glycosylation and methylation may also be prepared.

The present invention provides a novel method of obtaining both known and novel complex polyketides without the formation of mixtures of products differing only in having either an acetate or a propionate starter unit. In addition the present invention provides a method to obtain novel polyketides in which the starter

unit is an unusual starter unit which is derived by the action of a KSq domain on the enzyme-bound product of an AT of unusual specificity derived from an extension module of a natural Type I PKS. In particular the AT of extension module 4 of the FK506 PKS gene cluster preferentially incorporates an allyl side-chain; the AT of extension module 6 of the niddamycin PKS gene cluster preferentially incorporates a sidechain of structure $HOCH_2$-; and the ATs of extension module 5 of spiramycin and of extension module 5 of monensin incorporate an ethyl side chain. In each case the KSq domain is preferentially one that is naturally propionate-specific. Alternatively, any KS from an extension module of a Type I PKS may be converted into a KSq domain capable of decarboxylating a bound carboxylated acyl thioester, by site-directed mutagenesis of the active site cysteine residue to replace it by another residue, preferably glutamine. It is known that the animal fatty acid synthase, which shares many mechanistic features with Type I PKS, in the absence of acetyl-CoA, has a demonstrable malonyl-CoA decarboxylase activity (Kresze, G. B. *et al*. Eur. J. Biochem. (1977) 79:191-199). When treated with an alkylating agent such as iodoacetamide the fatty acid synthase is inactivated by specific modification of the active site cysteine of the KS, and

the resulting protein has an enhanced malonyl-CoA decarboxylase activity. The conversion of a fatty acid KS domain into a decarboxylase mirrors the genetically-determined change between the KS domains and the KSq domain in Type I PKSs. Indeed, the size and polarity characteristics of a glutamine side chain very closely approximate those of carboxamido-cysteine. The KSq to be used for decarboxylation of an unusual alkylmalonate unit is preferably selected from the same extension module of the same Type I PKS that provides the unusual AT, in order to optimise the decarboxylation of the unusual alkylmalonate, and the ACP to be used is preferably also the ACP of the same extension module.

Suitable plasmid vectors and genetically engineered cells suitable for expression of PKS genes incorporating an altered loading module are those described in PCT/GB97/01818 as being suitable for expression of hybrid PKS genes of Type I. Examples of effective hosts are *Saccharopolyspora erythraea*, *Streptomyces coelicolor*, *Streptomyces avermitilis*, *Streptomyces griseofuscus*, *Streptomyces cinnamonensis*, *Streptomyces fradiae*, *Streptomyces longisporoflavus*, *Streptomyces hygroscopicus*, *Micromonospora griseorubida*, *Streptomyces lasaliensis*, *Streptomyces venezuelae*, *Streptomyces antibioticus*, *Streptomyces lividans*, *Streptomyces*

rimosus, *Streptomyces albus*, *Amycolatopsis mediterranei*, and *Streptomyces tsukubaensis*. These include hosts in which SCP2*-derived plasmids are known to replicate autonomously, such as for example *S. coelicolor*, *S. avermitilis* and *S. griseofuscus*; and other hosts such as *Saccharopolyspora erythraea* in which SCP2*-derived plasmids become integrated into the chromosome through homologous recombination between sequences on the plasmid insert and on the chromosome; and all such vectors which are integratively transformed by suicide plasmid vectors.

Some embodiments of the invention will now be described with reference to the accompanying drawings in which:

Fig 1 is a diagram showing the functioning of 6-deoxyerythronolide B synthase (DEBS), a modular PKS producing 6-deoxyerythronolide B (6-DEB) a precursor of erythromycin A.

Fig 2 gives the amino acid sequence comparison of the KS domains and the CLF domains of representative Type II PKS gene clusters. The active site Cysteine (C) of the KS domains is arrowed in the Figure and aligns with the Glutamine (Q) or glutamic acid (E) of the CLF domains. The abbreviations used, and the relevant Genbank/EMBL accession numbers are: GRA: granaticin from *Streptomyces violaceoruber* (X63449); HIR: unknown

polyketide from *Saccharopolyspora hirsuta* (M98258); ACT, actinorhodin from *Streptomyces coelicolor* (X63449); CIN: unknown polyketide *from Streptomyces cinnamonensis* (Z11511); VNZ: jadomycin from *Streptomyces venezuelae* (L33245); NOG: anthracyclines from *Streptomyces nogalater* (Z48262); TCM: tetracenomycin from *S. glaucescens* (M80674); DAU: daunomycin from *Streptomyces sp.* C5 (L34880); PEU, doxorubicin from *Streptomyces peucetius* (L35560); WHI: WhiE spore pigment from *Streptomyces coelicolor* (X55942).

Fig 3 shows the gene organisation of the PKSs for three 16-membered ring macrolides, tylosin, spiramycin and niddamycin.

Fig 4 shows the amino acid sequence alignment of KSq-ATq loading didomains of the PKSs for niddamycin, platenolide(spiramycin), monensin, oleandomycin and tylosin. The sequences for the monensin and oleandomycin loading didomains have not been previously disclosed.

Fig. 5 The enyzmatic steps that convert 6-deoxyerythronolide B into erythromycin A *in Saccharopolyspora erythraea*

The present invention will now be illustrated, but is not intended to be limited, by means of some examples.

Example 1

Construction of the Recombinant Vector pPFL43

Plasmid pCJR24 was prepared as described in PCT/GB97/01819. pMD1TE is a pCJR24-based plasmid containing the gene encoding a hybrid polyketide synthase that contains the putative monensin PKS loading module (isolated from *S. cinnamonensis*) the DEBS extension modules 1 and 2 and the chain-terminating thioesterase. Plasmid pPFL43 was constructed as follows:

The following synthetic oligonucleotides: 5'-CCATATGGCCGCATCCGCGTCAGCGT-3' and 5'-GGCTAGCGGGTCCTCGTCCGTGCCGAGGTCA-3' are used to amplify the DNA encoding the putative monensin-producing loading module using a cosmid that contains the 5' end of the putative monensin-producing PKS genes from *S. cinnamonensis* or chromosomal DNA of *S. cinnamonensis* as template. The PCR product of 3.3 kbp is purified by gel electrophoresis, treated with T4 polynucleotide kinase and ligated to plasmid pUC18, which has been linearised by digestion with *Sma* I and then treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E.coli* DH10B cells and individual clones were checked for the desired plasmid pPFL40. Plasmid pPFL40 was identified by restriction pattern and sequence analysis.

Plasmid pPFL40 was digested with *Nde* I and *Nhe* I and

the 3.3 kbp fragment was purified by gel electrophoresis and ligated to ppND30-His previously digested with *Nde* I and *Nhe* I and treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E.coli* DH10B cells and individual clones were checked for the desired plasmid pPFL43. Plasmid pPFL43 was identified by restriction analysis.

Example 2

Construction of*S. erythraea* JC2/ pPFL43

Plasmid pPFL43 was used to transform *S.erythraea* JC2 protoplasts. The construction of strain JC2 from which the resident DEBS genes are substantially deleted is given in Pending Patent Application PCT/GB97/01818. Thiostrepton resistant colonies were selected in R2T20 medium containing 10 g/ml of thiostrepton. Several clones were tested for the presence of pPFL43 integrated into the chromosome by Southern blot hybridisation of their genomic DNA with DIG-labelled DNA containing the *mon* PKS fragment encoding for the loading module.

Example 3

Production of polyketides using *S. erythraea* JC2/pPFL43

A frozen suspension of strain *S. erythraea* JC2/ pPFL43 was inoculated in eryP medium, which has the following composition,

eryP medium

dextrose 50 g per l

nutrisoy flour 30 g per l

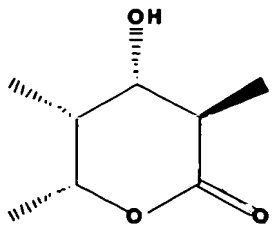
$(NH4)_2SO_4$ 3 g per l

$CaCO_3$ 6 g per l

NaCl 5 g per l

pH = 7.0

and containing 5 g/ml of thiostrepton. The inoculated culture was allowed to grow for seven days at 28-30°C. After this time the broth was filtered to remove mycelia and the pH adjusted to pH=3.0. The broth was extracted twice with two volumes of ethyl acetate and the combined extracts were washed with an equal volume of saturated sodium chloride, dried over anhydrous sodium sulphate, and the ethyl acetate was removed under reduced pressure, to give crude product. The product was shown to have the structure shown below, and by MS, GC-MS and ^1H NMR was found to be identical to an authentic sample.



Example 4

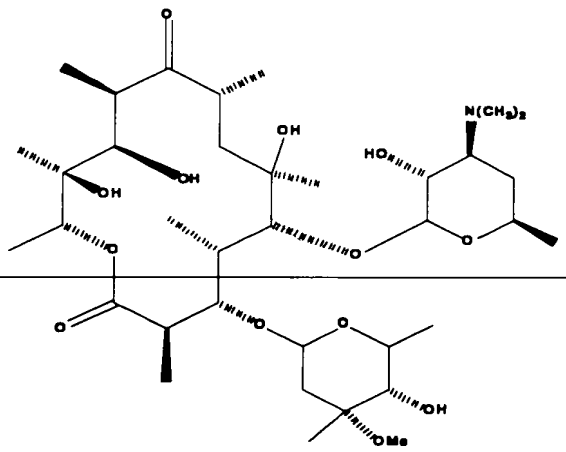
Construction of *S. erythraea* NRRL2338/pMD1TE

Plasmid pPFL43 was used to transform *S. erythraea* NRRL2338 protoplasts. Thiostrepton resistant colonies were selected in R2T20 medium containing 10 g/ml of thiostrepton. Several clones were tested for the presence of pPFL43 integrated into the chromosome by Southern blot hybridisation of their genomic DNA with DIG-labelled DNA containing the *mon* PKS fragment encoding for the loading module. A clone with an integrated copy of pPFL43 was selected in this way.

Example 5

Production of polyketides using *S. erythraea* NRRL 2338/pPFL43

A frozen suspension of strain *S. erythraea*/pMD1TE was used to inoculate eryP medium containing 5 g/ml of thiostrepton and allowed to grow for seven days at 28-30°C. After this time the broth was filtered to remove mycelia and the pH adjusted to pH=9. The supernatant was then extracted three times with an equal volume of ethyl acetate and the solvent was removed by evaporation. Products were analysed by HPLC/MS and one macrolide was identified with the structure shown, and identical by MS, GC-MS, and MS-MS with authentic material:



Example 6

Construction of the Recombinant Vector pPFL42

Plasmid pPFL42 is a pCJR24-based plasmid containing the gene encoding a hybrid polyketide synthase that contains the tylosin-producing PKS loading module, the erythromycin extension modules 1 and 2 and the chain-terminating thioesterase. Plasmid pPFL42 was constructed as follows:

The following synthetic oligonucleotides:

5'-CCATATGACCTCGAACACCGCTGCACAGAA-3' and

5'-GGCTAGCGGCTCCTGGGCTTCGAAGCTCTTCT-3'

were used to amplify the DNA encoding the tylosin-producing loading module using either cos6T (a cosmid that contains the tylosin-producing PKS genes from *S. fradiae*) or chromosomal DNA from *S. fradiae* as template. The PCR product of 3.3 kbp was purified by gel electrophoresis, treated with T4 polynucleotide kinase and ligated to plasmid pUC18, which had been linearised by digestion with *Sma* I and then treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E.coli* DH10B cells and individual clones were checked for the desired plasmid pPFL39. Plasmid pPFL39 was identified by restriction and sequence analysis.

Plasmid pPFL39 was digested with *Nde* I and *Nhe* I and the 3.3 kbp fragment was purified by gel electrophoresis and ligated to pND30 previously digested with *Nde* I and *Nhe* I and treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E.coli* DH10B cells and individual clones were checked for the desired plasmid pPFL42. Plasmid pPFL42 was identified by restriction analysis.

Example 7

Construction of *S. erythraea* JC2/pPFL42

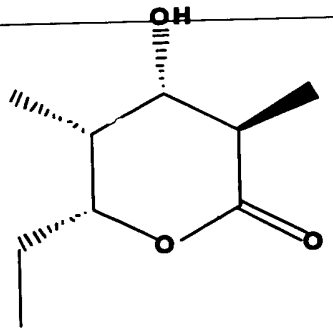
Plasmid pPFL42 was used to transform *S. erythraea* JC2 protoplasts. Thiostrepton resistant colonies were selected in R2T20 medium containing 10 g/ml of thiostrepton. Several clones were tested for the presence of pPFL42 integrated into the chromosome by Southern blot hybridisation of their genomic DNA with DIG-labelled DNA containing the tyl PKS fragment encoding for the loading module. A clone with an integrated copy of pPFL42 was identified in this way,

Example 8

Production of polyketides using *S. erythraea* NRRL 2338/pTD1TE

A frozen suspension of strain *S. erythraea* NRRL 2338/pPFL42 was used to inoculate eryP medium containing 5 g/ml of thiostrepton and allowed to grow for seven days at 28-30°C. After this time the broth was filtered to remove mycelia and the pH adjusted to pH=3. The broth was extracted twice with two volumes of ethyl acetate and the combined extracts were washed with an equal volume of saturated sodium chloride, dried over anhydrous sodium sulphate, and the ethyl acetate was removed under reduced pressure, to give crude product. The product was shown to have the structure shown below, and was identical, as

judged by MS, GC-MS, and ^1H NMR with an authentic sample:.



Example 9

Construction of *S. erythraea* NRRL2338/pPFL42

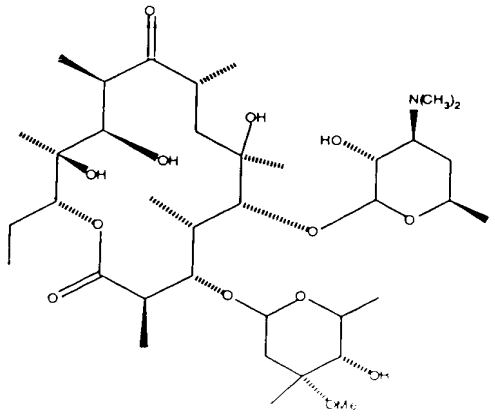
Plasmid pPFL42 was used to transform *S. erythraea* NRRL2338 protoplasts. Thiostrepton resistant colonies were selected in R2T20 medium containing 10 g/ml of thiostrepton. Several clones were tested for the presence of pPFL42 integrated into the chromosome by Southern blot hybridisation of their genomic DNA with DIG-labelled DNA

containing the *tyl* PKS fragment encoding for the loading module. A clone with an integrated copy of pPFL42 was identified in this way.

Example 10

Production of polyketides using *S. erythraea* NRRL2338/pPFL42

A frozen suspension of strain *S. erythraea* NRRL2338/pTD1TE was used to inoculate eryP medium containing 5 g/ml of thiostrepton and allowed to grow for seven days at 28-30°C. After this time the broth was filtered to remove mycelia and the pH adjusted to pH=9. The supernatant was then extracted three times with an equal volume of ethyl acetate and the solvent was removed by evaporation. Products were analysed by HPLC/MS and a macrolide was identified with the following structure, identical with that of authentic erythromycin A (together with other products, which were identified as the corresponding erythromycins B and D, the result of incomplete post-PKS processing):



Example 11

Construction of plasmid pPFL35

Plasmid pPFL35 is a pCJR24-based plasmid containing a PKS gene comprising a loading module, the first and second extension modules of DEBS and the chain terminating thioesterase. The loading module comprises the KSq domain DNA from the loading module of the oleandomycin PKS fused to the malonyl-CoA-specific AT of module 2 of the rapamycin PKS, in turn linked to the DEBS loading domain ACP. Plasmid pPFL35 was constructed via several intermediate plasmids as follows:

A 411 bp DNA segment of the *ery*AI gene from *S.erythraea* extending from nucleotide 1279 to nucleotide 1690 (Donadio, S. *et al.*, Science (1991) 2523:675-679) was amplified by PCR using the following synthetic oligonucleotide primers:-

5'-TGGACCGCCGCCAATTGCCTAGGCGGGCCGAACCCGGCT-3' and

5'-CCTGCAGGCCATCGCGACGACCGCGACCGGTTCGCC-3'

The DNA from a plasmid designated pKSW, derived from pT7-7 and DEBS1-TE in which new *Pst* I and *Hin*dIII sites had been introduced to flank the KS1 of the first extension module, was used as a template. The 441 bp PCR product was treated with T4 polynucleotide kinase and ligated to plasmid pUC18, which had been linearised by digestion with *Sma* I and then treated with alkaline phosphatase. The

ligation mixture was used to transform electrocompetent *E.coli* DH10B cells and individual clones were checked for the desired plasmid, pPFL26. The new *Mfe* I/*Avr* II sites bordering the insert are adjacent to the *Eco* RI site in the polylinker of pUC18. Plasmid pPFL26 was identified by restriction pattern and sequence analysis.

An *Mfe* I restriction site is located 112 bp from the 5' end of the DNA encoding the propionyl-CoA:ACP transferase of the loading module of DEBS. Plasmid pKSW was digested with *Mfe* I and *Pst* I and ligated with the 411 bp insert obtained by digesting plasmid pPFL26 with *Mfe* I and *Pst* I. The ligation mixture was used to transform electrocompetent *E.coli* DH10B cells and individual clones were checked for the desired plasmid, pPFL27. Plasmid pPFL27 contains a PKS gene comprising the DEBS loading module, the first and second extension modules of DEBS and the DEBS chain terminating thioesterase. Plasmid pPFL27 was identified by its restriction pattern.

Plasmid pPFL27 was digested with *Nde* I and *Avr* II and ligated to a 4.6kbp insert derived from digesting plasmid pMO6 (PCT/GB97/01818) with *Nde* I and *Avr* II. Plasmid pMO6 contains a PKS gene comprising the DEBS loading module, the first and second extension modules of DEBS and the DEBS chain terminating thioesterase, except that the DNA segment encoding the methylmalonate-specific AT within the first

extension module has been specifically substituted by the DNA encoding the malonate-specific AT of module 2 of the *rap* PKS. The ligation mixture was used to transform electrocompetent *E. coli* DH10B cells and individual clones were checked for the desired plasmid, pPFL28. Plasmid pPFL28 contains a hybrid PKS gene comprising the DEBS loading module, the malonate-specific AT of module 2 of the *rap* PKS, the ACP of the DEBS loading module, followed by the first and second extension modules of DEBS and the DEBS chain terminating thioesterase. Plasmid pPFL28 was identified by restriction analysis.

A DNA segment encoding the KSq domain from the *ole*AI gene of *S. antibioticus* extending from nucleotide 1671 to nucleotide 3385 was amplified by PCR using the following synthetic oligonucleotide primers:-

5'-CCACATATGCATGTCCCCGGCGAGGAA-3' and

5'-CCCTGTCCGGAGAAGAGGAAGGCGAGGCCG-3'

and chromosomal DNA from *Streptomyces antibioticus* as a template. The PCR product was treated with T4 polynucleotide kinase and ligated to plasmid pUC18, which had been linearised by digestion with *Sma* I and then treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E. coli* DH10B cells and individual clones were checked for the desired plasmid, pPFL31. The new *Nde* I site bordering the insert is adjacent

to the *Eco* RI site of the pUC18 polylinker while the new *Bsp* EI site borders the *Hin* dIII site of the linker region. Plasmid pPFL31 was identified by restriction and sequence analysis.

Plasmid pPFL31 was digested with *Nde* I and *Avr* II and the insert was ligated with plasmid pPFL28 that had been digested with *Nde* I and *Avr* II. The ligation mixture was used to transform electrocompetent *E.coli* DH10B cells and individual clones were checked for the desired plasmid, pPFL32. Plasmid pPFL32 was identified by restriction analysis.

Plasmid pPFL32 was digested with *Nde* I and *Xba* I and the insert was ligated to plasmid pCJR24, which had been digested with *Nde* I and *Xba* I and purified by gel electrophoresis. The ligation mixture was used to transform electrocompetent *E.coli* DH10B cells and individual clones were checked for the desired plasmid, pPFL35. Plasmid pPFL35 was identified by restriction analysis.

Example 12

Construction of *S. erythraea* JC2 / pPFL35

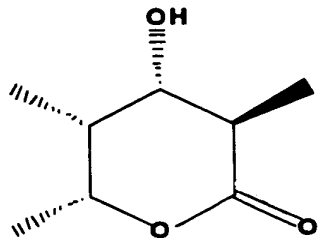
Plasmid pPFL35 was used to transform *S. erythraea* JC2 protoplasts. Thiostrepton resistant colonies were selected in R2T20 medium containing 10 g/ml of thiostrepton. Several clones were tested for the presence of pPFL35

integrated into the chromosome by Southern blot hybridisation of their genomic DNA with DIG-labelled DNA containing the *rap* PKS fragment encoding for module 2 acyltransferase. A clone with an integrated copy of pPFL35 was identified in this way.

Example 13

Production of polyketides using *S. erythraea* JC2 / pPFL35

A frozen suspension of strain *S. erythraea* JC2 / pPFL35 was used to inoculate eryP medium containing 5 g/ml of thiostrepton and allowed to grow for seven days at 28-30°C. After this time the broth was filtered to remove mycelia and the pH adjusted to pH=3. The broth was extracted twice with two volumes of ethyl acetate and the combined extracts were washed with an equal volume of saturated sodium chloride, dried over anhydrous sodium sulphate, and the ethyl acetate was removed under reduced pressure, to give crude product. The product was shown to have the following structure, and was found by MS, GC-MS and ^{1}H NMR to be identical to authentic material:



Example 14

Construction of *S. erythraea* NRRL2338/pPFL35

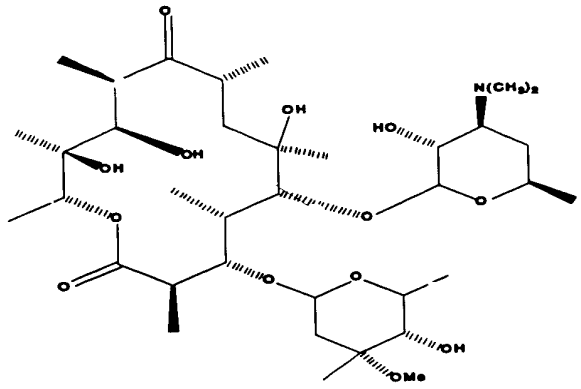
Plasmid pPFL35 was used to transform *S.erythraea* NRRL2338 protoplasts. Thiostrepton resistant colonies were selected in R2T20 medium (Yamamoto *et al.*) containing 10 g/ml of thiostrepton. Several clones were tested for the presence of pPFL35 integrated into the chromosome by Southern blot hybridisation of their genomic DNA with DIG-labelled DNA containing the *rap* PKS fragment encoding for module 2 AT. A clone with an integrated copy of pPFL35 was identified in this way.

Example 15

Production of polyketides using *S. erythraea* NRRL2338/pPFL35

A frozen suspension of strain *S. erythraea* NRRL2338/pPFL35 was used to inoculate eryP medium containing 5 g/ml of thiostrepton and allowed to grow for seven days at 28-30°C. After this time the broth was filtered to remove mycelia and the pH was adjusted to pH=9. The supernatant was then extracted three times with an equal volume of ethyl acetate and the solvent was removed by evaporation. Products were analysed by HPLC/MS and the following macrolide was identified, and its structure as

determined by MS and ^{1}H NMR was found to be identical to that of authentic material (it was accompanied by the corresponding erythromycins B and D, the products of incomplete processing by post-PKS enzymes:



Example 16

Construction of Recombinant Vector pPFL44

Plasmid pPFL44 is a pCJR24- based plasmid containing the gene encoding a hybrid polyketide synthase that contains the spiramycin PKS loading module, the erythromycin extension modules 1 and 2 and the chain-

terminating thioesterase. Plasmid pPFL44 was constructed as follows:

The following synthetic oligonucleotides:

5'-CCATATGTCTGGAGAACTCGCGATTTCCCGCAGT-3' and

5'-GGCTAGCGGGTCGTCGTCGTCCCGGCTG-3'

were used to amplify the DNA encoding the spiramycin-producing loading module using chromosomal DNA from the spiramycin producer *S. ambofaciens* prepared according to the method described by Hopwood *et al.* (1985). The PCR product of 3.3 kbp was purified by gel electrophoresis, treated with T4 polynucleotide kinase and ligated to plasmid pUC18, which had been linearised by digestion with *Sma* I and then treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E.coli* DH10B cells and individual clones were checked for the desired plasmid pPFL41. Plasmid pPFL41 was identified by restriction pattern and sequence analysis.

Plasmid pPFL41 was digested with *Nde* I and *Nhe* I and the 3.3 kbp fragment was purified by gel electrophoresis and ligated to pND30 (a plasmid derived from plasmid pCJR24 having as insert the *ave* PKS loading module and extension modules 1 and 2 or DEBS and the DEBS thioesterase) (PCTGB97/01810) previously digested with *Nde* I and *Nhe* I

and treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E.coli* DH10B cells and individual clones checked for the desired plasmid pPFL44. Plasmid pPFL44 was identified by restriction analysis.

Example 17

Construction of *S. erythraea* JC2/pSD1TE

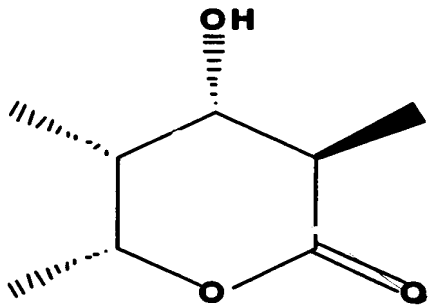
Plasmid pPFL44 was used to transform *S.erythraea* JC2 protoplasts. Thiostrepton resistant colonies were selected in R2T20 medium containing 10 g/ml of thiostrepton. Several clones were tested for the presence of pPFL44 integrated into the chromosome by Southern blot hybridisation of their genomic DNA with DIG-labelled DNA containing the srm PKS fragment encoding for the loading module. A clone with an integrated copy of pPFL44 was identified in this way.

Example 18

Production of polyketides using *S. erythraea* JC2/pPFL44

A frozen suspension of strain *S. erythraea* JC2/pPFL44 was used to inoculate eryP medium containing 5 g/ml of thiostrepton and allowed to grow for seven days at 28-30°C. After this time the broth was filtered to remove mycelia and the pH adjusted to pH=3. The broth was extracted twice

with two volumes of ethyl acetate and the combined extracts were washed with an equal volume of saturated sodium chloride, dried over anhydrous sodium sulphate, and the ethyl acetate was removed under reduced pressure, to give crude product. The product was shown to have the structure shown below and by GC-MS and ^{1}H NMR analysis was identical to authentic material:



Example 19

Construction of *S. erythraea* NRRL2338/pPFL44

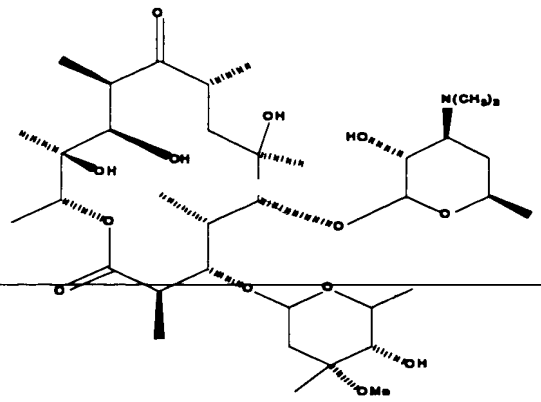
Plasmid pPFL44 was used to transform *S.erythraea* NRRL2338 protoplasts. Thiostrepton resistant colonies were

selected in R2T20 medium containing 10 g/ml of thiostrepton. Several clones were tested for the presence of pPFL44 integrated into the chromosome by Southern blot hybridisation of their genomic DNA with DIG-labelled DNA containing the spiramycin PKS fragment encoding for the loading module. A clone with an integrated copy of pPFL44 was grown was identified in this way.

Example 20

Production of polyketides using *S. erythraea* NRRL2338/pPFL44

A frozen suspension of strain*S. erythraea* NRRL2338/pPFL44 was used to inoculate eryP medium containing 5 g/ml of thiostrepton and allowed to grow for seven days at 28-30° C. After this time the broth was filtered to remove mycelia and the pH adjusted to pH=9. The supernatant was then extracted three times with an equal volume of ethyl acetate and the solvent was removed by evaporation. Products were analysed by HPLC/MS and a macrolide was identified which had the structure shown below, and as judged by MS and 1H NMR was identical to authentic material. :



Although the present invention is illustrated by the examples listed above, they should not be regarded as limiting the scope of the invention. The above descriptions illustrate for the first time the construction of a Type I PKS gene assembly containing a wholly or partly heterologous KSq-containing loading module and its use to obtain polyketide products of utility as synthetic intermediates or as bioactive materials such as antibiotics. It will readily occur to the person skilled in the art that a wholly or partly heterologous KSq-containing loading module from other PKS gene sets could be used to replace the loading module of DEBS, or indeed

into a quite different PKS gene assembly. It will also readily occur to the person skilled in the art that that the additional specificity provided by the more efficient discrimination made between methylmalonyl-CoA and malonyl-CoA by an ATq, followed by specific decarboxylation by a KSq, is preferable to the imperfect discrimination between propionyl-CoA and acetyl-CoA that is a feature of the DEBS loading module and of many other PKS loading modules, in that it maximises the production of a single product rather than a mixture differing from each other in the nature of the starter unit. The avoidance of such mixtures increases yields and avoids the need for tedious and difficult separation procedures.

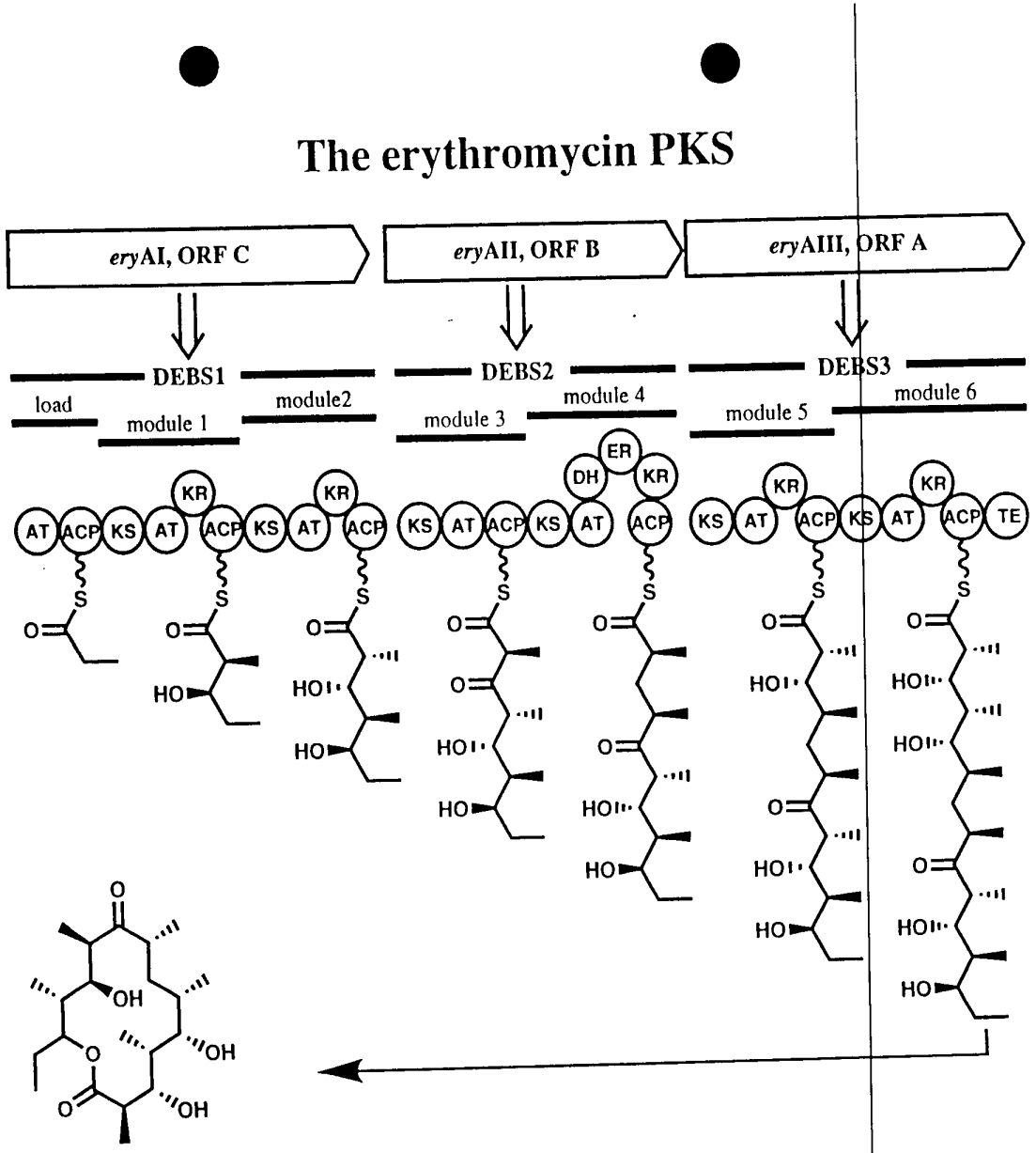

The erythromycin PKS
eryAI, ORF C
eryAII, ORF B
eryAIII, ORF A
DEBS1
DEBS2
DEBS3
load
module 1
module2
module 3
module 4
module 5
module 6
AT
ACP
KS
KR
DH
ER
TE
HO
OH
O
S

Fig. 1

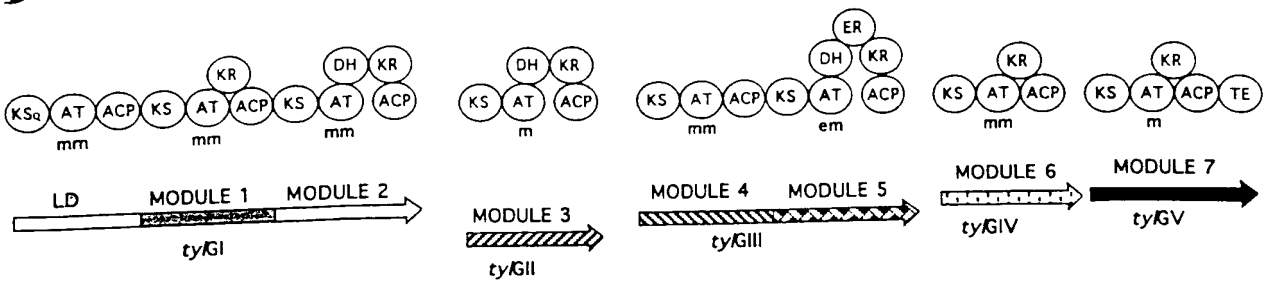
```
Se[illegible] Name: mole.bio.cam.ac.uk 3

KCLFDAU    --------------------------MVTGLGIVAPNGLGVGAIWDAVLNGRNGIGPLR
KCLFPEU    MTGTAARTASSQLHASPAGRRGLRGRAVVTGLGIVAPNGLGVGAYWDAVLNGRNGIGPLR
KCLFACT    -----------------------MSVLITGVGVVAPNGLGLAPYWSAVLDGRHGLGPVT
KCLFHIR    -----------------------MSTWVTGMGVVAPNGLGADDHWAATLKGRHGISRLS
KCLFGRA    -----------------MSTPDRRRAVVTGLSVAAPGGLGTERYWKSLLTGENGIAELS
KCLFNOG    ---------------------MTAAVVVTGLGVVAPTGLGVREHWSSTVRGASAIGPVT
KCLFTCM    -------------------MSAPAPVVVTGLGIVAPNGTGTEEYWAATLAGKSGIDVIQ
KCLFCIN    ---------------------MTP-VAVTGMGIAAPNGLGRPTTGRPPWAPRAASAAST
KCLFVNZ    ---------------------MSASVVVTGLGVAAPNGLGREDFWASTLGGKSGIGPLT
KCLFWHIE   ----------MSGPQRTGTGGGSRRAVVTGLGVLSPHGTGVEAHWKAVADGTSSLGPVT
KSGRA      ---------------------MTRRVVITGVGVRAPGGSGTKEFWDLLTAGRTATRPIS
KSHIR      ---------------------MTRRVVITGVGVRAPGGLGAKNFWELLTSGRTATRRIS
KSACT      ---------------------MKRRVVITGVGVRAPGGNGTRQFWELLTSGRTATRRIS
KSCIN      --------------------MTQRRVAITGIEVLAPGGLGRKEFWQLLSEGRTATRGIT
KSVNZ      --------------------MTARRVVITGIEVLAPGGTGSKAFWNLLSEGRTATRGIT
KSNOG      -----------------MKESINRRVVITGIGIVAPDATGVKPFWDLLTAGRTATRTIT
KSTCM      -----------------MTRHAEKRVVITGIGVRAPGGAGTAAFWDLLTAGRTATRTIS
KSDAU      ---------------------MNRRVVITGMGVVAPGAIGIKSFWELLLSGTTATRAIT
KSPEU      ---------------------MNRRIVITGIGVVAPGAVGTKPFWELLLSGTTATRAIS
KSWHI      --------------------MTRRRVAVTGIGVVAPGGIGTPQFWRLLSEGRTATRRIS
                                  :**:  :  :*  .  *                .

KCLFDAU    RFADDGRLGRLAGEVSDFVP-EDHLPKRLLVQTDPMTQMTALAAAEWALREAGCAPSS--
KCLFPEU    RFTGDGRLGRLAGEVSDFVP-EDHLPKRLLAQTDPMTQY-ALAAAEWALRESGCSPSS--
KCLFACT    RFDVSRYPATLAGQIDDFHA-PDHIPGRLLPQTDPSTRL-ALTAADWALQDAKADPES-L
KCLFHIR    RFDPTGYPAELAGQVLDFDA-TEHLPKRLLPQTDVSTRF-ALAAAAWALADAEVDPAE-L
KCLFGRA    RFDASRYPSRLAGQIDDFEA-SEHLPSRLLPQTDVSTRY-ALAAADWALADAGVGPESGL
KCLFNOG    RFDAGRYPSKLAGEVPGFVP-EDHLPSRLMPQTDHMTRL-ALVAADWAFQDAAVDPSK-L
KCLFTCM    RFDPHGYPVRVGGEVLAFDA-AAHLPGRLLPQTDRMTQH-ALVAAEWALADAGLEPEK-Q
KCLFCIN    RFDPSGYPAQLAGEIPGFRA-AEHLPGRLVPQTDRVTRL-SLAAADWALADAGVEVAA-F
KCLFVNZ    RFDPTGYPARLAGEVPGFAA-EEHLPSRLLPQTDRMTRL-ALVAADWALADAGVRPEE-Q
KCLFWHIE   REGCAHLPLRVAGEVHGFDA-AETVEDRFLVQTDRFTHF-ALSATQHALADARFGRADVD
KSGRA      FFDASPFRSRIAGEI-DFDAVAEGFSPREVRRMDRATQF-AVACTRDALADSGLDTGA-L
KSHIR      FFDPTPNRSQIAAEC-DFDPEHEGLSPREIRRMDRAAQF-AVVCTRDAVADSGLEFEQ-V
KSACT      FFDPSPYRSQVAAEA-DFDPVAEGFGPRELDRMDRASQF-AVACAREAFAASGLDPDT-L
KSCIN      FFDPAPFRSKVAAEA-DFCGLENGLSPQEVRRMDRAAQF-AVVTAR-AVEDSGAELAA-H
KSVNZ      FFDPTPFRSRVAAEI-DFDPEAHGLSPQEIRRMDRAAQF-AVVAAR-AVADSGIDLAA-H
KSNOG      AFDPSPFRSRIAAEC-DFDPLAEGLTPQQIRRMDRATQF-AVVSARESLEDSGLDLGA-L
KSTCM      LFDAAPYRSRIAGEI-DFDPIGEGLSPRQASTYDRATQL-AVVCAREALKDSGLDPAA-V
KSDAU      TFDATPFRSRIAAEC-DFDPVAAGLSAEQARRLDRAGQF-ALVAGQEALTDSGLRIGE-D
KSPEU      TFDATPFRSRIAAEC-DFDPVAAGLSAEQARRLDRAGQF-ALVAGQEALADSGLRIDE-D
KSWHI      LFDPSGLRSQIAAEC-DFEPSDHGLGLATAQRCDRYVQF-ALVAASEAVRDANLDMNR-E
                 :..:    *     .           *   :  ::      :.  :

KCLFDAU    -PLEAGVITASASGGFASGQRELQNLWSKG-----PAHVSAYMSFAWFY-AVNTGQIAIR
KCLFPEU    -PLEAGVITASASGGFAFGQRELQNLWSKG-----PAHVSAYMSFAWFY-AVNTGQIAIR
KCLFACT    TDYDMGVVTANACGGFDFTHREFRKLWSEG-----PKSVSVYESFAWFY-AVNTGQISIR
KCLFHIR    PEYGTGVITSNATGGFEFTHREFRKLWAQG-----PEFVSVYESFAWFY-AVNTGQISIR
KCLFGRA    DDYDLGVVTSTAQGGFDFTHREFHKLWSQG-----PAYVSVYESFAWFY-AVNTGQISIR
KCLFNOG    PEYGVGVVTASSAGGFEFGHRELQNLWSLG-----PQYVSAYQSFAWFY-AVNTGQVSIR
KCLFTCM    DEYGLGVLTAAGAGGFEFGQREMQKLWGTG-----PERVSAYQSFAWFY-AVNTGQISIR
KCLFCIN    DPLDMGVVTASHAGGFEFGQDELQKLLGQG-----QPVLSAYQSFAWFY-AVNSGQISIR
KCLFVNZ    DDFDMGVVTASASGGFEFGQGELQKLWSQG-----SQYVSAYQSFAWFY-AVNSGQISIR
KCLFWHIE   SPYSVGVVTAAGSGGGEFGQRELQNLWGHG-----SRHVGPYQSIAWFY-AASTGQVSIR
KSGRA      DPSRIGVALGSAVASATSLENEYLVMSDSGREWLVDPAHLSPMMFDYLSPGVMPAEVAWA
KSHIR      PPERIGVSLGSAVAAATSLEQEYLVLSDGGREWQVDPAYLSAHMFDYLSPGVMPAEVAWT
KSACT      DPARVGVSLGSAVAAATSLEREYLLLSDSGRDWEVDAAWLSRHMFDYLVPSVMPAEVAWA
KSCIN      PPHRIGVVVGSAVGATMGLDNEYRVVSDGGRLDLVDHRYAVPHLYNYLVPSSFAAEVAWA
KSVNZ      DPYRVGVTVGSAVGATMGLDEEYRVVSDGGRLDLVDHAYAVPHLYDYMVPSSFSAEVAWA
KSNOG      DASRTGVVVGSAVGCTTSLEEEYAVVSDSGRNWLVDDGYAVPHLFDYFVPSSIAAEVAHD
KSTCM      NPERIGVSIGTAVGCTTGLDREYARVSEGGSRWLVDHTLAVEQLFDYFVPTSICREVAWE
KSDAU      SAHRVGVCVGTAVGCTQKLESEYVALSAGGANWVVDPHRGAPELYDYFVPSSLAAEVAWL
KSPEU      SAHRVGVCVGTAVGCTQKLESEYVALSAGGAHWVVDPGRGSPELYDYFVPSSLAAEVAWL
KSWHI      DPWRAGATLGTAVGGTTRLEHDYVLVSERGSRWDVDDRRSEPHLERAFTPATLSSAVAEE
             *.  .        . :   :   *                     :        ::

KCLFDAU    -HDLRGPVGVVVAEQAGGLDALAHAR-RKVRGGAE-LIVSGAMDSSLCP-YGMAAQVRSG
KCLFPEU    -HDLRGPVGVVVAEQAGGLDALAHAR-RKVRGGAE-LIVSGAVDSSLCP-YGMAAQVKSG
KCLFACT    -HGMRGPSSALVAEQAGGLDALGHAR-RTIRRGTP-LVVSGGVDSALDP-WGWVSQIASG
KCLFHIR    -HGLRGPGSVLVAEQAGGLDAVGHGG--AVRNGTP-MVVTGGVDSSFDP-WGWVSHVSSG
KCLFGRA    -NTMRGPSAALVGEQAGGLDAIGHAR-RTVRRGPG-WCSAVASTRRSTR-GASSSQLSGG
KCLFNOG    -HGLRGPGGVLVTEQAGGLDALGQAR-RQLRRGLP-MVVAGAVDGSPCP-WGWVAQLSSG
KCLFTCM    -HGMRGHSSVFVTEQAGGLDAAAHAA-RLLRKGTLNTALTGGCEASLCP-WGLVAQIPSG
KCLFCIN    -HGMKGPSGVVVSDQAGGLDALAQAR-RLVRKGTP-LIVCGAVEPRSAPGAGSPSSPAGG
KCLFVNZ    -NGMKGPSGVVVSDQAGGLDAVAQAR-RQIRKGTR-LIVSGGVDASLCP-WGWVAHVASD
KCLFWHIE   -NDFKGPCGVVAADEAGGLDALAHAA-LAVRNGTD-TVVCGATEAPLAP-YSIVCQLGYP
KSGRA      -AGAEGPVTMVSDGCTSGLDSVGYAV-QGTREGSADVVVAGAADTPVSPIVVACFDAIKA
KSHIR      -VGAEGPVAMVSDGCTSGLDSLSHAC-SLIAEGTTDVMVAGAADTPITPIVVSCFDAIKA
KSACT      -VGAEGPVTMVSTGCTSGLDSVGNAV-RAIEEGSADVMFAGAADTPITPIVVACFDAIRA
KSCIN      -VGAEGPSTVVSTGCTSGIDAVGIAV-ELVREGSVDVMVAGAVDAPISPIP-CVLDAIKA
KSVNZ      -VGAEGPNTVVSTGCTSGLDSVGYARGELIREGSADVMIAGSSDAPISPITMACFDAIKA
KSNOG      RIGAEGPVSLVSTGCTSGLDAVGRAA-DLIAEGAADVMLAGATEAPISPITVACFDAIKA
KSTCM      -AGAEGPVTVVSTGCTSGLDAVGYGT-ELIRDGRADVVVCGATDAPISPITVACFDAIKA
KSDAU      -AGAEGPVNIVSAGCTSGIDSIGYAC-ELIREGTVDVMLAGGVDAPIAPITVACFDAIRV
KSPEU      -AGAEGPVNIVSAGCTSGIDSIGYAC-ELIREGTVDAMVAGGVDAPIAPITVACFDAIRA
KSWHI      -FGVRGPVQTVSTGCTSGLDAVGYAY-HAVAEGRVDVCLAGAADSPISPITMACFDAIKA
             .*    .      :.*:*:  .  .               *            .

KCLFDAU    RLSGSDDPTAGYLPFDRRAAGHVPGEG-GAILAVEDAERVAERG-GKVYGSIAGT-ASFD
KCLFPEU    RLSGSDNPTAGYLPFDRRAAGHVPGEG-GAILTVEDAERAAERG-AKVYGSIAGYGASFD
KCLFACT    RISTATDPDRAYLPFDERAAGYVPGEG-GAILVLEDSAAAEARGRHDAYGELAGCASTFD
KCLFHIR    RVSRATDPGRAYLPFDVAANGYVPGEG-GAILLLEDAESAKARG-ATGYGEIAGYAATFD
KCLFGRA    LVSTVADPERAYLPFDVDASGYVPGEG-GAVLIVEDADSARARG---AERIYVRSPLRRD
KCLFNOG    GLSTSDDPRRAYLPFDAAAGGHVPGEG-GALLVLESDESARARGVTRWYGRIDGYAATFD
KCLFTCM    FLSEATDPHDAYLPFDARAAGYVPGEG-GAMLVAERADSARERDAATVYGRIAGHASTFD
KCLFCIN    -MSDSDEPNRAYLPFDRDGRGYVPGGGRGVVPPLERAEAAPARG-AEVYGE-AGPLARL-
KCLFVNZ    RLSTSEEPARGYLPFDREAQGHVPGEG-GAILVMEAAEAARERG-ARIYGEIAGYGSTFD
KCLFWHIE   ELSRATEPDRAYRPFTEAACGFAPAEG-GAVLVVEEEAAARERG-ADVRATVAGHAATFT
```

Session Name: mole.bio.cam.ac.uk

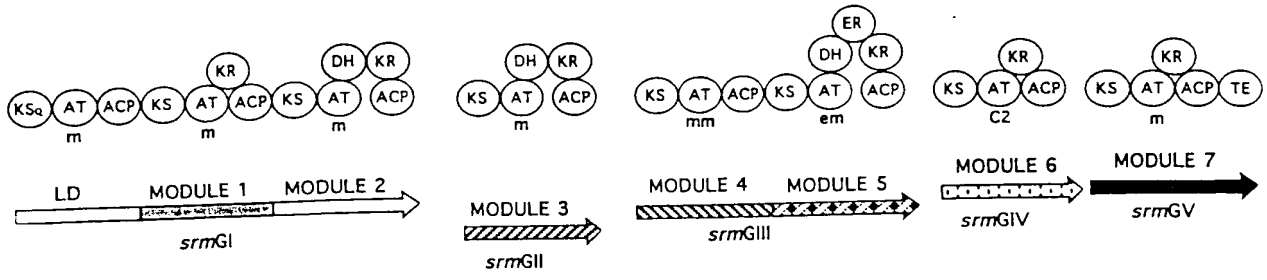
```
[illegible]GRA          TTPRNDDPAHASRPFDGTRNGFVLAEG-AAMFVLEEYEAAQRRG-AHIYAEVGGYATRSQ
[illegible]HIR          TTPRNDDPEHASRPFDNSRNGFVLAEG-AALFVLEELEHARARG-AHVYAEISGCATRLN
KSACT           TTARNDDPEHASRPFDGTRDGFVLAEG-AAMFVLEDYDSALARG-ARIHAEISGYATRCN
KSCIN           TTPRHDAPATASRPFDSTRNGFVLGEG-AAFFVLEELHSARRRG-AHIYAEIAGYATRSN
KSVN[illegible]         TTNRYDDPAHASRPFDGTRNGFVLGEG-AAVFVLEELESARARG-AHIYAEIAGYATRSN
KSNO[illegible]         TTPRNDTPAEASRPFDRTRNGFVLGEG-AAVFVLEEFEHARRRG-ALVYAEIAGFATRCN
KSTCM           TSANNDDPAHASRPFDRNRDGFVLGEG-SAVFVLEELSAARRRG-AHAYAEVRGFATRSN
KSDAU           TSDHNDTPETLA-PFSRSRNGFVLGEG-GAIVVLEEAEAAVRRG-ARIYAEIGGYASRGN
KSPEU           TSDHNDTPETASRPFSRSRNGFVLGEG-GAIVVLEEAEAAVRRG-ARIYAEIGGYASRGN
KSWHI           TSPNNDDPAHASRPFDADRNGFVMGEG-AAVLVLEDLEHARARG-ADVYCEVSGYATFGN

KCLFDAU         -PPPGSGRP---SALARAVETALADAGLDRSDIAVVFADGAA-VGELDVAEAEALASVFG
KCLFPEU         -PPPGSGRP---SALARAVETALADAGLDGSDIAVVFADGAA-VPELDAAEAEALASVFG
KCLFACT         -PAPGSGRP---AGLERAIRLALNDAGTGPEDVDVVFADGAG-VPELDAAEARAIGRVFG
KCLFHIR         -PAPGSERP---PALRRAIELALADAELRPEQVDVVFADAAG-VAELDAIEAAAIRELFG
KCLFGRA         -PAPGSGRP---PALGRAAELALAEAGLTPADISVVFADGAG-VPELDRAEADTLARLFG
KCLFNOG         -PPPGSGRP---PNLLRAAQAALDDAEVGPEAVDVVFADASG-TPDEDAAEADAVRRLFG
KCLFTCM         -ARPGTGRP---TGPARAIRLALEEARVAPEDVDVVYADAAG-VPALDRAEAEALAEVFG
KCLFCIN         -PAPHSGRG---STRAHAIRTALDDAGTAPGDIRRVFADGGGRYPN-DRAEAEAISEVFG
KCLFVNZ         -PRPGSGRE---PGLRKAIELALADAGAAPGDIDVVFADAAA-VPELDRVEAEALNAVFG
KCLFWHIE        GAGRWAESR---EGLARAIQGALAEAGCRPEEVDVVFADALG-VPEADRAEALALADALG
KSGRA           -AYHMTGLKKDGREMAESIRAALDEARLDRTAVDYVNAHGSG-TKQNDRHETAAFKRSLG
KSHIR           -AYHMTGLKTDGREMAEAIRVALDLARIDPTDIDYINAHGSG-TRQNDRHETAAYKRALG
KSACT           -AYHMTGLKADGREMAETIRVALDESRTDATDIDYINAHGSG-TKQNDRHETAAFKKALG
KSCIN           -AYHMTGLR-DGAEMAEAIRLALDEARLNPEQVDYINAHGSG-TKQNDRHETAAFKKALG
KSVNZ           -AYHMTGLRPDGAEMAEAIRVALDEARMNPTEIDYINAHGSG-TRQNDRHETAAFKRSLG
KSNOG           -AFHMTGLRPDGREMAEAIGVALAQAGKAPADVDYVNAHGSG-TRQNDRHETAAFKRSLG
KSTCM           -AFHMTGLKPDGREMAEAITAALDQARRTGDDLHYINAHGSG-TRQNDRHETAAFKRSLG
KSDAU           -AYHMTGLRADGAEMAAAITAALDEARRDPSDVDYVNAHGTA-TRQNDRHETSAFKRSLG
KSPEU           -AYHMTGLRADGAEMAAAITAALDEARRDPSDVDYVNAHGTA-TKQNDRHETSAFKRSLG
KSWHI           -AYHMTGLTKEGLEMARAIDTALDMAELDGSAIDYVNAHGSG-TQQNDRHETAAVKRSLG

[illegible]CLFDAU         P--HRVPVTVPKTLTGRLYSGAGPLDVATGLLALRDEVVPATGHVH-PDPDLPLDVVTGR
KCLFPEU         P--RRVPVTVPKTLTGRLYSGAGPLDVATALLALRDEVVPATAHVD-PDPDLPLDVVTGR
KCLFACT         R--EGVPVTVPKTTTGRLYSGGGPLDVVTALMSLREGVIAPTAGVTSVPREYGIDLVLGE
KCLFHIR         P--SGVPVTAPKTMTGRLYSGGGPLDLVAALLAIRDGVIPPTVHTAEPVPEHQLDLVTGD
KCLFGRA         P--RGVPVTAPKALTGRLCAGGGPADLAAALLALRDQVIPATGRHRAVPDAYALDLVTGR
KCLFNOG         P--YGVPVTAPKTMTGRLSAGGAALDVATALLALREGVVPPTVNVSRPRPEYELDLVLA-
KCLFTCM         P--GAVPVTAPKTMTGRLYAGGAALDVATALLSIRDCVVPPTVGTGAPAPGLGIDLVLHQ
KCLFCIN         P--GRVPVTCPRTMTGRLHSGAAPLDVACALLAMRAGVIPPTVHID-PCPEYDLDLVLYQ
KCLFVNZ         T--GAVPVTAPKTMTGRLYSGAAPLDLAAAFLAMDEGVIPPTVNVE-PDAAYGLDLVVGG
KCLFWHIE .      PHAARVPVTAPKTGTGRAYCAAPVLDVATAVLAMEHGLIPPTPHVL--DVCHDLDLVTGR
KSGRA           EHAYAVPVSSIKSMGGHSLGAIGSIEIAASVLAIEHNVVPPTANLHTPDPECDLDYVPLT
KSHIR           EHAYRTPVSSIKSMVGHSLGAIGSIEVAACALAIEHGVVPPTANLHEPDPECDLDYVPLT
KSACT           EHARRTPVSSIKSMVGHSLGAIGSLEIAACVLALEHGVVPPTANLRTSDPECDLDYVPLE
KSCIN           EHAYRTPVSSIKSMVGHSLGAIGSIEIAASALAMEYDVVPPTANLHTPDPECDLDYVPLT
KSVNZ           DHAYRTPVSSIKSMVGHSLGAIGSIEIAASALAMEHNVVPPTGNLHTPDPECDLDYVR-S
KSNOG           DHAYRVPVSSIKSMIGHSLGAIGSLEIAASVLAITHDVVPPTANLHEPDPECDLDYVPLR
KSTCM           DHAYRVPISSVKSMIGHSLGAAGSLEVAATALAVEYGAIPPTANLHDPDPELDLDYVPLT
KSDAU           QRAYDVPVSSIKSMIGHSLGAIGSLELAACALAIEHGVIPPTANYEEPDPECDLDYVPNV
KSPEU           DHAYRVPISSIKSMIGHSLGAVGSLEVAATALAVEYGVIPPTANLHDPDPELDLDYVPLT
KSWHI           EHAYRVPISSIKSMIGHSLGAVGSLEVAATALAVEYGVIPPTANLHDPDPELDLDYVPLT
                EHAYATPMSSIKSMVGHSLGAIGSIELAACVLAMAHQVVPPTANYTTPDPECDLDYVPRE

KCLFDAU         PRAMADARAALVVARGHGGFNSALVVRGAA-------
KCLFPEU         PRSLADARAALLVARGYGGFNSALVVRGAA-------
KCLFACT         PRSTAPRTA-LVLARGRWGFNSAAVLRRFAPTP----
KCLFHIR         PRHQQLGTA-LVLARGKWGFNSAVVVRGVTG------
KCLFGRA         PREAALSAA-LVLARGRHGFNSAVVVTLRGSDHRRPT
KCLFNOG         PRRTPLARA-LVLARGRGGFNAAMVVAGPRAETR---
[illegible]LFTCM         PRELRVDTA-LVVARGMGGFNSALVVRRHG-------
[illegible]LFCIN         VRPAALRTA-LGGARGHGGFNSALVVRAGQ-------
KCLFVNZ         PRTAEVNTA-LVIARGHGGFNSAMVVRSAN-------
KCLFWHIE        ARPAEPRTA-LVLARGLMGSNSALVLRRGAVPPEGR-
KSGRA           AREQRVDTV-LTVGSGFGGFQSAMVLHRPEEAA-----
KSHIR           AREQRVDTV-LSVGSGFGGFQSAMVLRRLGGANS---
KSACT           ARERKLRSV-LTVGSGFGGFQSAMVLRDAETAGAAA-
KSCIN           ARDQRVDSV-LTVGSGFGGFQSAMVLTSAQ---RSTV
KSVNZ           CREQLTDSV-LTVGSGFGGFQSAMVLARPE---RKIA
KSNOG           ARACPVDTV-LTVGSGFGGFQSAMVLCGPGSRGRSAA
KSTCM           AREQRVDTV-LSVGSGFGGFQSAAVLARPKETRS---
KSDAU           AREKRVRHA-LTVGSGFGGFQSAMLLSRPER------
KSPEU           AREKRVRHA-LTVGSGFGGFQSAMLLSRLER------
KSWHI           ARERTLRHV-LSVGSGFGGFQSAVVLSGSEGGLR---
```

mole:~/ks2%

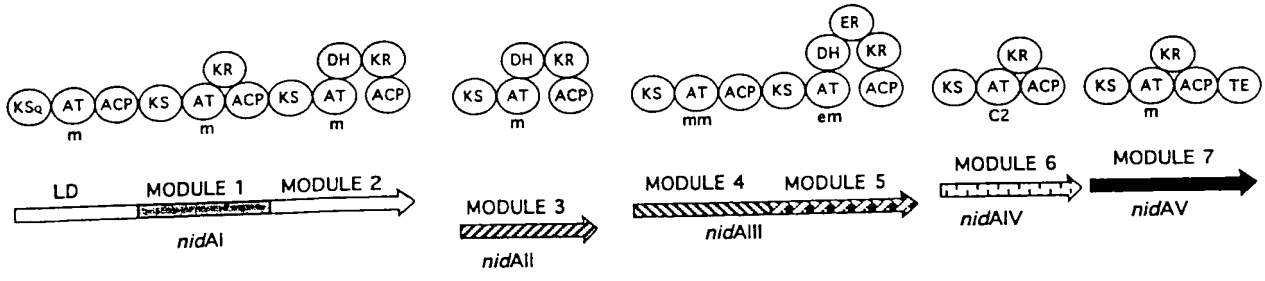
RGANISATION OF THE TYLOSIN-PRODUCING POLYKETIDE SYNTHASE



ORGANISATION OF THE SPIRAMYCIN-PRODUCING POLYKETIDE SYNTHASE



ORGANISATION OF THE NIDDAMYCIN-PRODUCING POLYKETIDE SYNTHASE



m: malonyl transferase
mm: methylmalonyl transferase
em: ethylmalonyl transferase
C2: unknown C2 unit transferase

Fig 3

Fig. 4 (1 of 3) Alignment of KSq-ATq loading modules of modular polyketide synthases

```
                                                                      50
           1
niddamycin ~~~~~~~~~~ ~~~~~~~~~~ MAGHGDATAQ KAQDAEKSED GSDAIAVIGM
platenolide ~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~MS GELAISRSDD RSDAVAVVGM
monensin   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~MAAS ASASPSGPSA GPDPIAVVGM
oleandomycin~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~MHVPGEE NGHSIAIVGI
tylosin    MSSALRRAVQ SNCGYGDLMT SNTAAQNTGD QEDVDGPDST HGGEIAVVGM

                                                                     100
           51
niddam...  SCRFPGAPGT AEFWQLLSSG ADAVVTAADG RRR..... .....GTIDA
platenol.  ACRFPGAPGI AEFWKLLTDG RDAIGRDADG RRR..... .....GMIEA
monensin   ACRLPGAPDP DAFWRLLSEG RSAVSTAPPE RRRADSGLHG P...GGYLDR
oleandom   ACRLPGSATP QEFWRLLADS ADALDEPPAG RFPTGSLSSP PAPRGGFLDS
tylosin    SCRLPGAAGV EEFWELLRSG RGMPTRQDDG TWRAA..... .......LED

                                                                     150
           101
niddam...  PADFDAAFFG MSPREAAATD PQQRLVLELG WEALEDAGIV PESLRGEAAS
platenol.  PGDFDAAFFG MSPREAAETD PQQRLMLELG WEALEDAGIV PGSLRGEAVG
monensin   IDGFDADFFH ISPREAVAMD PQQRLLLELS WEALEDAGIR PPTLARSRTG
oleandom   IDTFDADFFN ISPREAGVLD PQQRLALELG WEALEDAGIV PRHLRGTRTS
tylosin    HAGFDAGFFG MNARQAAATD PQHRLMLELG WEALEDAGIV PGDLTGTDTG

                                                                     200
           151
niddam...  VFVGAMNDDY ATLLH.RAGA PTDTYTATGL QHSMIANRLS YFLGLRGPSL
platenol.  VFVGAMHDDY ATLLH.RAGA PVGPHTATGL QRAMLANRLS YVLGTRGPSL
monensin   VFVGAFWDDY TDVLNLRAPG AVTRHTMTGV HRSILANRIS YAYHLAGPSL
oleandom   VFMGAMWDDY AHLAHARGEA ALTRHSLTGT HRGMIANRLS YALGLQGPSL
tylosin    VFAGVASDDY A.VLTRRSAV SAGGYTATGL HRALAANRLS HFLGLRGPSL

                                                                     250
           201
niddam...  VVDTGQSSSL VAVALAVESL RGGTSGIALA GGVNLVLAEE GS.AAMERVG
platenol.  AVDTAQSSSL VAVALAVESL RAGTSRVAVA GGVNLVLADE GT.AAMERLG
monensin   TVDTAQSSSL VAVHLACESI RSGDSDIAFA GGVNLICSPR TTELAAARFG
oleandom   TVDTGQSSSL AAVHMACESL ARGESDLALV GGVNLVLDPA GT.TGVERFG
tylosin    VVDSAQSASL VAVQLACESL RRGETSLAVA GGVNLILTEE ST.TVMERMG

                                                                     300
           251
niddam...  ALSPDGRCHT FDARANGYVR GEGGAIVVLK PLADALADGD RVYCVVRGVA
platenol.  ALSPDGRCHT FDARANGYVR GEGGAAVVLK PLADALADGD PVYCVVRGVA
monensin   GLSAAGRCHT FDARADGFVR GEGGGLVVLK PLAAARRDGD TVYCVIRGSA
oleandom   ALSPDGRCYT FDSRANGYAR GEGGVVVVLK PTHRALADGD TVYCEILGSA
tylosin    ALSPDGRCHT FDARANGYVR GEGGGAVVLK PLDAALADGD RVYCVIKGGA

                                                                     350
           301
niddam...  TGNDGGGPGL TVPDRAGQEA VLRAACDQAG VRPADVRFVE LHGTGTPAGD
platenol.  VGNDGGGPGL TAPDREGQEA VLRAACAQAR VDPAEVRFVE LHGTGTPVGD
monensin   VNSDGTTDGI TLPSGQAQQD VVRLACRRAR ITPDQVQYVE LHGTGTPVGD
oleandom   LNNDGATEGL TVPSARAQAD VLRQAWERAR VAPTDVQYVE LHGTGTPAGD
tylosin    VNNDGGGASL TTPDREAQEA VLRQAYRRAG VSTGAVRYVE LHGTGTRAGD
```

```
                                                                  400
            351
niddam...   PVEAEALGAV YGTGRP..AN EPLLVGSVKT NIGHLEGAAG IAGFVKAALC
platenol.   PVEAHALGAV HGSGRP..AD DPLLVGSVKT NIGHLEGAAG IAGLVKAALC
monensin    PIEAAALGAA LGQDAA..RA VPLAVGSAKT NVGHLEAAAG IVGLLKTALS
oleandom    PVEAEGLGTA LGTARP..AE APLLVGSVKT NIGHLEGAAG IAGLLKTVLS
tylosin     PVEAAALGAV LGAGADSGRS TPLAVGSVKT NVGHLEGAAG IVGLIKATLC

                                                                  450
            401
niddam...   LHERALPASL NFETPNPAIP LERLRLKVQT AHAALQPGTG GGPLLAGVSA
platenol.   LRERTLPGSL NFATPSPAIP LDQLRLKVQT AAAELPLAPG GAPLLAGVSS
monensin    IHHRRLAPSL NFTTPNPAIP LADLGLTVQQ DLADWP..RP EQPLIAGVSS
oleandom    IKNRHLPASL NFTSPNPRID LDALRLRVHT AYGPWP..SP DRPLVAGVSS
tylosin     VRKGELVPSL NFSTPNPDIP LDDLRLRVQT ERQEW.NEED DRPRVAGVSS

            451                                                   500
niddam...   FGMGGTNCHV VLEETPGG.. .......... .......... ...RQPAE.T
platenol.   FGIGGTNCHV VLEHLPSR.. .......... .......... ...PTPAV.S
monensin    FGMGGTNGHV VVA....AAP DSVAVPEPVG VPERVEVPEP VVVSEPVVVP
oleandom    FGMGGTNCHV VLSELRNAGG DGAGKGPYTG TEDRLGATEA EKRPDPATGN
tylosin     FGMGGTNVHL VIAEAPAAAG SSGAGGSGAG SGAGISAVSG VV........

                                                                  550
            501
niddam...   GQADACLFSA SPMLLLSARS EQALRAQAAR LREHL..EDS GADPLDIAYS
platenol.   VAAS...LPD VPPLLLSARS EGALRAQAVR LGETV..ERV GADPRDVAYS
monensin    TPWP...... .....VSAHS ASALRAQAGR LRTHLAAHRP TPDAARVGHA
oleandom    GPDPAQDTHR YPALILSARS DAALRAQAER LRHHL.EHSP GQRLRDTAYS
tylosin     .......... ..PVVVSGRS RVVVREAAGR LAE..VVEAG GVGLADVAVT

                                                                  600
            551
niddam...   LATTRTRFEH RAAVPCGDPD RLSSALAALA AGQTPRGVRI GS..TDADGR
platenol.   LASTRTLFEH RAVVPCGGRG ELVAALGGFA AGRVSGGVRS GR..A.VPGG
monensin    LATTRAPLAH RAVLLGGDTA ELLGSLDALA EGAETASIVR GEAYT..EGR
oleandom    LATRRQVFER HAVVTGHDRE DLLNGLRDLE NGLPAPQVLL GRTPTPEPGG
tylosin     MAD.RSRFGY RAVVLARGEA ELAGRLRALA GGDPDAGVVT G...AVLDGG

                                                                  650
            601
niddam...   LALLFTGQGA QHPGMGQELY TTDPHFAAAL DEVCEELQRC GTQNLREVMF
platenol.   VGVLFTGQGA QWVGMGRGLY AGGGVFAEVL DEVLSMVGEV DGRSLRDVMF
monensin    TAFLFSGQGA QRLGMGRELY AVFPVFADAL DEAFAALDVH LDRPLREIVL
oleandom    LAFLFSGQGS QQPGMGKRLH QVFPGFRDAL DEVCAELDTH LGRLL.....
tylosin     VVVGAAPGGA GAAGGAGAAG GAGGGGVVLV FPGQGTQWVG MGAGLLGSSE

                                                                  700
            651
niddam...   TPDQPD.... .......... .......... LLDRTEYTQP ALFALQTALY
platenol.   GDVDVDAGAG ADAGAGAGAG VGSGSGSVGG LLGRTEFAQP ALFALEVALF
monensin    GETDSGGNVS GENVIGEGA. ......DHQA LLDQTAYTQP ALFAIETSLY
oleandom    .GPEAGPPLR DVMFAERGT. ......AHSA LLSETHYTQA ALFALETALF
tylosin     VFAASMRECA RALSVHVGWD LLEVVSGGAG .LERVDVVQP VTWAVMVSLA

                                                                  750
            701
niddam...   RTLTARGTQA HLVLGHSVGE ITAAHIAGVL DLPDAARLIT ARAHVMGQLP
platenol.   RALEARGVEV SVVLGHSVGE VAAATVAGVL SLGDAVRLVV ARGGLMGGLP
monensin    RLAASFGLKP DYVLGHSVGE IAAAHVAGVL SLPDASALVA TRGRLMQAVR
oleandom    RLLVQWGLKP DHLAGHSVGE IAAAHAAGIL DLSDAAELVA TRGALMRSLP
tylosin     RYWQAMGVDV AAVVGHSQGE IAAATVAGAL SLEDAAAVVA LRAGLIGRYL
```

Fig. 4 continued 2 of 3

```
                751                                                  800
niddam...   HG.GAMLSVQ AAEHDLDQLA HTHG..VEIA AVNGPTHCVL SGPRTALEET
platenol.   VG.GGMWSVG ASESVVRGVV EGLGEWVSVA AVNGPRSVVL SGDVGVLESV
monensin    AP.GAMAAWQ ATADEAAEQL AGHERHVTVA AVNGPDSVVV SGDRATVDEL
oleandom    GG.GVMLSVQ APESEVAPLL LGREAHVGLA AVNGPDAVVV SGERGHVAAI
tylosin     AGRGAMAAVP LPAGEVEAGL .AKWPGVEVA AVNGPASTVV SGDRRAVAGY

                801                                                  850
niddam...   AQHLREQNVR HTWLKVSHAF HSALMDPMLG AFRDTLNTLN Y..QPPTIPL
platenol.   VASLMGDGVE YRRLDVSHGF HSVLMEPVLG EFRGVVESLE FGRVRPGVVV
monensin    TAAWRGRGRK AHHLKVSHAF HSPHMDPILD ELRAVAAGLT FHE..PVIPV
oleandom    EQILRDRGRK SRYLRVSHAF HSPLMEPVLE EFAEAVAGLT FRA..PTTPL
tylosin     VAVCQAEGVQ ARLIPVDYAS HSRHVEDLKG ELERVLSGI. .RPRSPRVPV

                851                                                  900
niddam...   ISNLTGQIA. .....DPNHL CTPDYWIDHA RHTVRFADAV QTAHHQGTTT
platenol.   VSGVSGGVV. .....GSGEL GDPGYWVRHA REAVRFADGV GVVRGLGVGT
monensin    VSNVTGELVT ATATGSGAGQ ADPEYWARHA REPVRFLSGV RGLCERGVTT
oleandom    VSNLTG.... ..APVDDRTM ATPAYWVRHV REAVRFGDGI RALGKLGTGS
tylosin     CSTVAGEQPG EPVF...... .DAGYWFRNL RNRVEFSAVV GGLLEEGHRR

                901                                                  950
niddam...   YLEIGPHPTL TTLLHHTL.. .DNP..... .........T TIPTLHRERP
platenol.   LVEVGPHGVL TGMAGECLGA GDDV..... .........V VVPAMRRGRA
monensin    FVELGPDAPL SAMARDCFPA P......... .ADRSRPRPA AIATCRRGRD
oleandom    FLEVGPDGVL TAMARACVTA APEPGHRGEQ GADADAHTAL LLPALRRGRD
tylosin     FIEVSAHPVL V......... .....HAIEQ TAEAADRSVH ATGTLRRQDD

                951
niddam...   EPETLTQAIA AVGVRTDGID WAVLCGASRP RRVELPTYAF
platenol.   EREVFEAALA TVFTRDAGLD ATALHTGSTG RRIDLPTTPF
monensin    EVATFLRSLA QAYVRGADVD FTRAYGATAT RRFPLPTYPF
oleandom    EARSLTEAVA RLHLHGVPMD WTSVLGGDVS .RVPLPTYAF
tylosin     SPHRLLTSTA EAWAHGATLT WDPAL..PPG HLTTLPTYPF
```

niddam: niddamycin; platenol: platenolide I (spiramycin); oleandom: oleandomycin.

FIG. 4 (continued)(3 of 3)

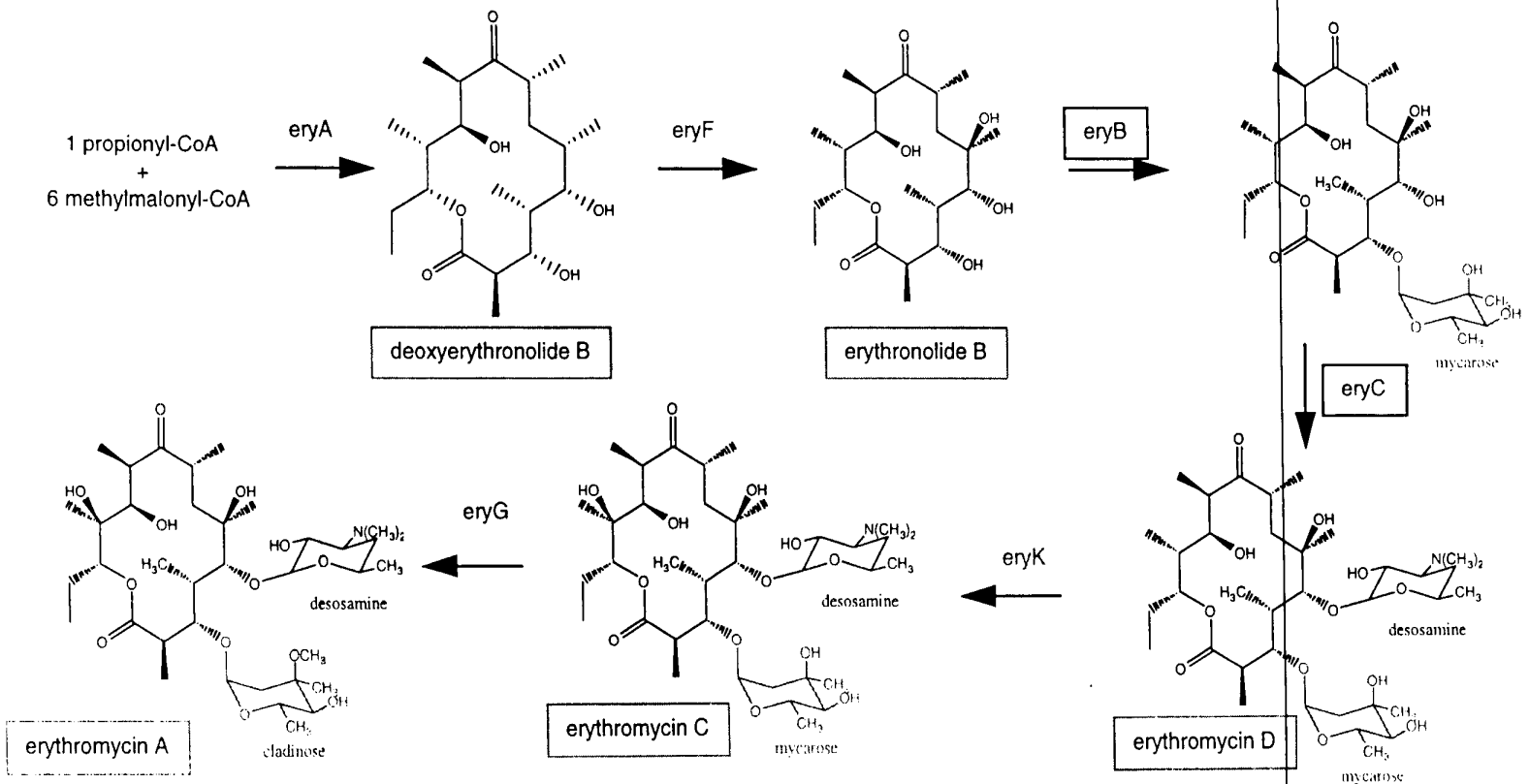

1 propionyl-CoA
+
6 methylmalonyl-CoA
eryA
deoxyerythronolide B
eryF
erythronolide B
eryB
mycarose
eryC
erythromycin D
desosamine
mycarose
eryK
erythromycin C
desosamine
mycarose
eryG
erythromycin A
desosamine
cladinose

Fig. 5